CHAPTER 3

FORMULATING THE PURPOSE STATEMENT/RESEARCH QUESTION

Definition of a Purpose Statement

Usually the research problem formulated in the form of a question at the initial stage helps the researcher what to focus on throughout the research. Here are some examples of research questions:

- Is there a relationship between the introversion-extroversion tendencies of Turkish students at secondary schools and their proficiency in English as a foreign language?
- What type of study skills do Freshman students at Çukurova University need?
- How well textbooks on Business English reflect the use of English used by the people involved in business?
- How do teachers at YADIM (Center for Foreign Languages) of Çukurova University feel about the new syllabus on listening and speaking designed for this year?

Aside from a research question the purpose of the research can be given in a statement:

- The purpose of this study is to compare the effectiveness of two writing methods, free and guided, in increasing the writing performance of university students at the intensive English Language Programs.
- This paper reviews the influence of the communicative task on curriculum development at English Language Schools.
- This empirical investigation seeks to determine the attitudes of both L1 and L2 listeners toward specific regional accents of American English.

As for the common characteristics that entitle them to be research questions or purpose statements in each case,

- there is a need to find some information.
- answers (at least in part) can be found by means of collecting data from available sources.
- data can be collected by means of a particular technique: observation, interview, questionnaire, tests etc.
- the population to be examined is described.
- the sample drawn out of the population and the variables are explicitly defined. This way the topic is narrowed down and the scope of the study is limited to a specific group.
- there are two or more variables that seem to affect one another.

- there is an implication of indicating the relation(s) or the difference(s) between (among) these variables.
- the topic is researchable and/or has an empirical verification. For example, a question like " What is the meaning of *vital*?," does not require much except for looking the word up in the dictionary, and this is something everybody can do. Therefore, the topic cannot be considered for research.
- terms specific to the field are defined or explicitly stated.

Here is another research question:

- Should driving be taught at high schools?

This question cannot be searched for an answer because it does not have any empirical referents. In other words, there is no scientific way of finding an answer to a question that requires a stable outcome with the insertion of the modal *should*. If this question is modified as indicated below, it becomes researchable because under these conditions, a researcher can interview people to find out their opinions.

- Do you think driving should be taught at high schools? or
- What do people think about including driving into the high school curriculum?

Formulation of a Purpose Statement

In order to be able to formulate a good research question or a purpose statement, the terms *variable population* and *sample* need to be defined. The word 'variable' is derived from the verb 'vary' and refers to the variation of any class of objects. Objects are referred to in relation to their variation of color, speed, size and function. For instance, round tables differ from rectangular tables, or dining tables from tennis tables. Thus, tables vary according to their shape or to their function, and accordingly each of these groups qualifies the class as a variable. Due to their different shapes and functions, these variables may have different effects on the behaviour of people (the population). Since the research cannot be conducted on all the people carrying the similar characteristics, a relatively small group of people representing the whole population (sample) is chosen for research purposes. "Researchers choose certain variables to investigate because they have a suspicion that these variables are somewhat related and that if they can discover the nature of this relationship, it can help us make more sense of the world in which we live" (Fraenkel & Wallen, 1990, p. 36).

Since the task of stating expected effect of the variables on the population is not easy, the way (a) the population, (b) the sample, (c) the variables are selected should be explicitly defined along with (d) the specific terms to be used in the study. Additionally, what have been stated about variables and population, need to be consistently cited in other sections of the research.

Population

In choosing the population the researcher should see that individuals included in the population have the same characteristics except the one(s) stated in the variables. In other words, the other variables should be kept constant in order to arrive at a sound conclusion. In a particular research, the population might be kept the same in reference to age, sex, educational background depending on the purpose of the research. In other words, populations may be any *size*, and may cover any *attribute* and any *geographical* area. What is important, however, is that the generalizations arrived at the end of the study should be limited to the chosen population. For instance, the results of a study conducted at one school in Ankara cannot be generalized for the school system in that city. For that reason, "the key is to define your population in sufficient detail so that others may determine how applicable your finding might be to other situations" (Gay, 1987, p. 103)

In order to make a generalization about the whole population, samples representing different sections of the population need to be included. In most cases, the population to derive generalizations from is hardly available. Therefore, the researcher realistically selects the population from what is available.

For instance, if the interest is to find out about the writing skills of children at the elementary school, then the population should comprise the elementary school children. This characteristic is too broad for the study to be conducted; therefore, the population should be narrowed down by adding extra characteristics that would limit the scope of the study (see also Chapter 2). For this purpose, another type of limitation such as focusing only on one grade in the elementary school can be set. This might still be broad. Then, the population can be narrowed down by considering all the sixth-graders (a) in the country, (a) in a particular geographical region, (c) in a city, (d) in a district in the city, (e) in two elementary schools. This indicates that when the *target population* (see Example 1) cannot be reached at, attempt is made to collect data from the *accessible population* (see Example 2). When the accessible population is found to be too broad; then, the researcher chooses a *sample* (see Example 3) among the accessible population.

Examples:

1) Target population:	All the sixth-graders in Turkey.
2) Accessible population:	All the sixth-graders in Adana.
3) Sample:	300 fifth-graders selected from different elementary schools in Adana.

Sampling

After having decided how many subjects should be included in the sample, effort needs to be made to select these subjects so that they would accurately represent the population characteristics. There are different methods of random sampling. The most commonly cited ones are simple random sampling, stratified random sampling, cluster random sampling, systematic random sampling. *In simple random sampling, each individual in the defined population has an equal change of being selected* but as a result of sheer luck or as a result of some systematic screening, these individuals are not included in the sample. The subjects for the study, for example, could be selected by drawing out their names from a box as in the lottery system (*simple random sampling*) or by picking up every tenth name from a list (*systematic random sampling*). In both instances, the

subjects are chosen individually to form the sample. In some cases, selection of sampling is based on groups rather than on individuals. For instance, subjects belonging to the same group (e.g. the whole class) are chosen all together to form part or the whole of a sample. This type of sampling is called *cluster sampling* (Fraenkel & Wallen, 1990, p. 74).

If the focus of the research is to *compare* two groups of people with distinct characteristics, then *stratified random sampling* is used. For instance, if the learning rate between girls and boys in the first grade is investigated, after the population is defined, the subjects are grouped separately in two samples (1. boys, 2. girls). The size for each sample group is based on whether *proportional stratified sampling* or *equal size stratified sampling* is applied. In proportional stratified sampling, the number in each sample (boys/girls) should be proportional to the total number of boys and girls in the defined population. For example, if there are 300 boys and 400 girls in the defined population, and the 10 percent of the population is decided to be included in the sample, then 30 boys and 40 girls are randomly selected.

In equal size sampling, ignoring the proportion of the boys and girls within the population, equal number of subjects are chosen from both sexes. For instance, 30 boys and 30 girls would be selected for the same type of study.

What makes the research reliable is the way the samples are selected. Therefore, the steps to be followed for each type of sampling are indicated below:

Steps in simple random sampling

Aside from the lottery system based on complete chance, simple random selection can be applied using a table of random numbers. The procedure provided by Gay (1981) is given below exemplifying each step :

1. Identify and define the population.
   [e.g. : All the High School English Teachers in Adana (800 teachers)]
2. Determine the desired sample size.
   [e.g. : 10% of these teachers (80 teachers)]
3. List all members of the population
   [e.g. : You get the lists from the directors of the schools]
4. Assign all individuals on the list a consecutive number from zero to the required number
   [e.g. : Numbering the teachers on the list from 000 to 800 or 00 to 89]
5. Select an arbitrary number in the table of random numbers. (Close your eyes and point!)
   [e.g. : Out of the following numbers , you underline the one that is chosen:
   96455
   76054
   63**276**
   12869
   76107
   67532, etc.]

6 For the selected number, look at only the appropriate number of digits. For example, if a population has 800 members, you only need to use the

last 3 digits of the number; if a population has 90 members you only need to use the last 2 digits.
[e.g. : Since the size of the sample is determined to be 800, only the last three digits of these random numbers are taken into consideration (e.g. 276).]

7. If the number corresponds to the number assigned to any of the individuals in the population, then that individual is in the sample. For example, if a population had 500 members and the number selected was 375, the individual assigned 375 would be in the sample; if a population had only 300 members, then 375 would be ignored.
[e.g. : In this case, if the number selected is 576, it is selected but if it is 801 or above, it is ignored]
8. Go to the next number in the column and repeat step 7.
[e.g. : The next number in the column is 12869. The last three digits is over 800; therefore, it is ignored.]
9. Repeat step 8 until the desired number of individuals has been selected for the sample.
[e.g. : The above steps are applied to 76107, 67532, and since the last three digits for these two numbers are 107, 532, teachers that have been assigned these numbers are included into the sample. The same procedure is applied to the rest of the numbers in the column till 80 subjects are selected.] (p. 105)

Steps in systematic random sampling

The first three steps are the same as in the simple random sampling:

1. Identify and define the population.
2. Determine the desired sample size.
3. Obtain a list of the population.
4. Decide what K is equal to by dividing the size of the population by the desired sample size.
[e.g. : Since the defined population is 800 teachers and the size of the sample is 80, 800 is divided by 80 and the K for this sampling is determined to be 10
(K= total number of the defined population/total number of the defined sample).
5. Start at some random place at the top of the population list.
6. Starting at that point, take every Kth name on the list until the desired sample is reached.
[e.g. : Starting from the randomly selected name at the top of the list, every tenth name is included in the sample. For example, if the name randomly selected at the top of the list was the fourth name, then 14th, 24th, 34th, 44th, 54th names and the consequent names that fall on the frequency count of 10 will be included in the sample.] (p. 113)

Steps in stratified random sampling

Following are the steps adopted from Gay (1981):

1. Identify and define the population.
2. Determine the desired sample size.
3. Identify the subgroups according to the variables to be looked into. For instance, the 800 teachers in the defined population can be divided into the years of training they have received at tertiary level. Then teachers could be divided into two: (a) teachers with two years of training, (b) teachers with more than two years of training. Another variable that would lead to a division would be the years of experience the teachers have: (a) teachers with teaching experience between 1-5 years, (b) teachers with teaching experience between 6-10 years, and so on.
4. Classify all members of the population according to the identified subgroups. For instance, out of 800 teachers, there would be 100 teachers with two years of training and 700 hundred with four years of training.
5. Using a table of random numbers, select a sample for each group. The number of teachers may vary depending on what type of stratified sampling is going to be adopted. If a proportional representation is desired, then 10 among teachers with two years of training and 70 among teachers with four years of training would be randomly selected. If an equal sized sample is desired, then 40 from each group are selected to form the sample.

Steps in cluster random sampling

In cluster random sampling the subjects are chosen in groups rather than as individuals.

1. Identify and define the population. For instance, your population is defined as 800 freshman students studying English at public schools in Adana.
2. Determine the desired sample size. For instance, there will be 80 students in the sample.
3. Define your cluster. In other words, will a cluster comprise of students in a class; teachers in a school; people belonging to a certain club, etc.? According to the population defined in Step 1, the logical cluster would be a school.
4. Obtain a list of clusters that comprise the defined population. In other words, find out how many clusters there are in the defined population. For instance, first there is a need to know how many schools there are in Adana.
5. Examine the number of subjects in each cluster and accordingly estimate how many subjects there will be in each cluster. For instance, the estimated number of freshman students in each school is 40.
6. Decide how many clusters will be needed by taking into consideration the number of the subjects to be included in the sample. For instance,

the sample is going to be composed of 80 subjects and there are 20 high schools in Adana. Since the study aims at taking 10 percent of the population as a sample, then, two schools need to be randomly chosen.

7. Select the number of clusters using a table of random numbers. In the selection of the sample, make sure that all the population members are well represented by the members of the clusters to be chosen. For instance, in the selection, attention has to be paid to include schools from different social backgrounds to have an accurate representation of the defined population

In some cases, researchers are obliged to choose their subjects from their nearby environment due to some inconvenience. In other cases, depending on their judgement, they attempt to choose their subjects according to the purpose of their study. Both of these procedures yield a *biased sampling* (Moore, 1983, p. 124)/*nonrandom sampling* (Fraenkel & Wallen, 1990, p. 76).

Variables

Definition of variables

In describing the population, the variables should be determined and defined very carefully. Following is a list of some variables considered useful in defining the characteristics of the population or the sample:

age, sex, ethnicity, marital status, socioeconomic status, grade level, occupation, level of education, level of income, geographic region, interest groups, academic or nonacademic status, level of intelligence, level of achievement in different skills, cognitive development level, language proficiency, L1 or L2 membership.

The terms used in defining the specifications of the variables should not be too general or too specific. Moore (1983, p. 47) demonstrates different levels of specificity of the statement of the purpose. For instance, while the term "anxiety" is too general, "self-report inventory of anxious situations while taking a test" is too specific. Moore suggests that, at a general conceptual level, the use of "test anxiety" is more appropriate. The same reasoning is valid for defining the population. Just mentioning about "students" is too general. "Fifth-grade students" seems to be at a general conceptual level but "fifth-grade students at Urla Elementary School" would be too specific.

Types of variables

There are different ways of categorizing variables:

- *Quantitative* as opposed to *categorical*
- *Manipulated* as opposed to *outcome*
- *Dependent* as opposed to *independent*

Quantitative variables are expressed in degrees and amounts but not in kind. Therefore, the questionnaire designed for investigation comprises alternative responses including expressions that indicate the beginning and the end of a continuum. For instance, a question, such as "How much homework does your teacher give you over the semester?" may have several options varying from "none" to "too much." If such a question is asked, the students answering the questionnaire may choose one of the following options:

A. None B. Very little C. Enough D. Much E. Too much

If the purpose is to compare the amount of homework given in two different classes, the options might be limited to "less" and "more." This is because here, the focus is not on the relative opinion of the individual members of the population but on the overall situation within two classes and especially the difference between these two classes as far as homework assignment is considered. Thus, the students are expected to express their opinion about the amount of homework given to them by choosing the term either *more* or *less***.**

*Categorical variable*s do not vary in quantity. Instead, they differ qualitatively referring to the treatment and the methods applied.

For example:

- If the purpose of the research is to investigate the effectiveness of Method X on reading achievement, Method X is the *categorical*, and the reading achievement is the *quantitative variable*.

This is because methods can be categorized, but reading achievement can only be expressed in terms of quantity. This quantity can be expressed either in numbers (e.g. 100) or letters (e.g. A) to indicate the highest quantity. These numbers, however, do not have any arithmetical values although they can be used in the calculation of frequency counts as seen in the following examples:

- If the purpose of the research is to investigate the relation of gender and the major of the students at a university, both the gender and the major of the students are *categorical variables*. While a male/female notation can be used in recording this information, it is customary to assign arbitrary numbers to each possible category. For instance, if the variable is gender, the symbol 1 might represent male and 2 female. These numbers, used for the purpose of classification and categorization, do not have any arithmetical value. In other words, they can not be added or multiplied. If all the 1's/ 2's are counted in reference to the major of the students, a frequency count of how many males/females there are in that major is obtained.

- If the purpose of the study is to investigate the relation between the reading achievement of the students and their motivation, both reading achievement and motivation are *quantitative* variables because they can be measured not in terms of numbers but in terms of degrees and/or amounts.

Manipulated variables are also called *treatment variables* or *experimental variables* because the researcher manipulates, in other words, sets up a condition to

serve the purpose of the research and applies it as a treatment on the subjects (the people the researcher is examining or experimenting with). The effects of the treatment are the *outcome variables* as seen in the following purpose statement:

> The main purpose of this study is to determine the effectiveness of team work as compared to no teamwork in increasing the level of self-confidence in first-grade children.

Here, the experiment is conducted at least with two different classes. The manipulation is the application of teamwork. This is applied only to one class. The expected effect of the treatment is that the level of self confidence will increase. For that reason, the effect of the treatment is called the outcome variable.

In some instances, manipulated variables correspond to *independent variables* and the outcome variables correspond to *dependent variables*. This is because an independent variable is "presumed to affect or influence other variables. A dependent variable (or outcome variable) is presumed to be affected by one or more independent variables" (Fraenkel & Wallen 1990, p. 43). This distinction can be clarified by the following example:

- If the purpose of the research is to examine the effect of newly designed structural syllabus on the development of learners' reading comprehension, structural syllabus is the *manipulated/independent* variable, and the degree of improvement in learners' reading comprehension is the *outcome/dependent* variable.

However, independent and dependent variables vary from manipulated and outcome variables from different perspectives. For instance, it is possible to investigate more than one independent variable affecting a dependent variable as seen in the following example:

- If the purpose of the study is to investigate how English native speakers and Japanese native speakers at professional graduate, and undergraduate levels evaluate and edit ESL (English as a Second Language) compositions written by Japanese college students, English native speakers/Japanese native speakers and professional graduate/undergraduate levels are two *independent* variables, and the *dependent* variable is the degree of evaluation of the ESL compositions.

Another point that the independent variables differ from manipulated variables is that, although manipulated variables are controlled, all the independent variables cannot be controlled. Under these circumstances, these variables are called *extraneous variables*. An extraneous variable is not a variable of the central concern in a particular research. However, it might affect the outcome in an undesired manner. When interpreting the findings, therefore, results should not be generalized beyond the groups included in the study. The personality, the intelligence of the person, past events are among these extraneous variables because these are independent variables that have unintended effects on a dependent variable.

Moore (1983) gives the following striking example to demonstrate the importance of extraneous variables:

> ... suppose we are interested in determining the effect of three counseling techniques __ rational-emotive, gestalt, and no counseling __ in lowering the anxiety levels of volunteer clients at a local mental health facility. However, we neglect to screen our subjects for severe psychopathology. Instead of having just average neurotics in the counseling group, a paranoid psychotic is included in the rational-emotive group. At the conclusion of the study, anxiety level scores are significantly higher in a posttest of that group when compared to the others. (p. 141)

As seen in the above example, the fact that one of the subjects being paranoid, and influencing both the dependent and the independent variable, can be indicated as an extraneous variable since including a subject of this type into the study group was not controlled at the initial stage of the research. In other words, it was not known that this specific person was paranoid. In most cases, within these unknown situations there will always be some extraneous variables effecting the result of the study without even being detected.

Definition of Terms Specific to the Field

The terms specific to the field may not be familiar to the reader and thus need to be clearly defined. Fraenkel and Wallen (1990, p. 24) mention three ways of clarifying terms in a research question:

One way is providing a constitutive definition which is referred to as the *dictionary approach*. In this technique, sometimes the terms are not explicitly defined, or other technical terms are used within the definition. Therefore, researchers do not seem to favor this technique of defining terms.

A second way is an *example*. This might be efficient in some cases but the reader who is not familiar with the term will not most probably be aware of the unit exemplifying this term. For instance, in defining the term "humanistic classroom", how much it will help the reader if the name of the institution where this technique is applied is given.

A third method of clarification used in defining important terms is *operational definition.* By operational definitions, a researcher specifies the actions or operations necessary to measure or identify a specific term used in the research. For example, for the term " humanistic classroom,"
Fraenkel and Wallen (1990) give the following operational definitions:

1. Any classroom *identified* by specific experts as constituting an example of a humanistic classroom.
2. Any classroom *judged* (by an observer spending at least one day per week for four to five weeks) to possess all the following characteristics:

a. no more than three children working with the same materials at the same time;
b. the teacher never spending more than twenty minutes per day addressing the class as a group;
c. at least half of every class period open for students to work on projects of their own choosing at their own pace;

d. several (more than three) sets of different kinds of educational materials available for every student in the class to use;
e. nontraditional-type seating- students sit in circles, small grouping of seats, or even on the floor to work on their projects;
f. frequent (at least two per week) discussions of ideas in which students are encouraged to give their viewpoints about topics being read about in their textbooks (p. 25).

In addition to the above issues regarding the formulation of a research question or a purpose statement, Fraenkel and Wallen (1990, p. 31), underline the importance of feasibility, clarity and significance, and ethical value in the fulfilment of a purpose statement:

- If the question stated in the research cannot be investigated within a reasonable amount of time, energy and money, it will not be feasible.
- If the variables and the population under concern are not stated clearly with no operational definitions given, the readers will not be able to see the aim and the main frame of the research.
- If the designed research does not provide contribution to the advancement of knowledge in the field , if it does not promise to improve the human condition or the existing situation in the field, the research cannot be considered significant.
- If the result of the research has the risk of bringing physical or psychological harm to anyone, it is not considered ethical either. All subjects in the research should be assured that the data collected from them or about them will be kept confidential.

In the formulation of the purpose statement, special attention should be given for not being biased. For that reason, instead of the verb phrases (e.g. to show, to prove, to demonstrate and to confirm) that reflect some forms of a bias, neutral expressions (e.g. to determine, to compare, to investigate, to differentiate) should be used in order to avoid making value judgements. If any professional predictions need to be made due to a supportive theory or personal experience, this needs to be indicated in the hypothesis. Hypotheses are the only appropriate places to formulate such predictions.