ISSN 0258-2279 LITERATOR 15 No. 3 Nov. 1994

The interpretation and evaluation of quantitative research studies in Second Language Acquisition

Carisma Dreyer and Johann L. van der Walt
Department of English Language and Literature
Potchefstroom University for CHE
POTCHEFSTROOM

Abstract

This article argues that many quantitative research studies are deficient as far as their method of research is concerned. The 'Method' section should be critically evaluated in order to determine whether the findings that are reported are valid and reliable. Each of the sub-sections usually included in the method section is discussed, and a checklist for the evaluation of quantitative research studies is provided.

1. Introduction

In carrying out research, the issue or question one wants to investigate should form the point of departure in deciding on the appropriate research method. Many issues in second language acquisition are appropriately researched by means of a quantitative research method. This method has become very popular in recent years, because it is a powerful research tool which allows researchers to go beyond the identification and linear description of language learning phenomena and to draw formal inferences from the data about expected frequencies of occurrence, to assess the likelihood that phenomena are generalizable beyond a given instance, or to compare adequacies of existing theories and models to account for the phenomena in question.

Many quantitative studies, however, are deficient and reveal many weaknesses. Henning (1986), for example, points out that many studies do not provide any estimate of the validity and reliability of the instrumentation and procedures used to elicit data. In this regard Wolfson (1986:690) states that "No matter what else we do, we must remember that if data are inadequate, there is always the danger that the theory and conclusions drawn from them could be unreliable and misleading". Researchers wishing to replicate studies fall into the trap of using inadequate test instruments or tools for analyses. Many post-graduate students are unable to analyze and evaluate quantitative studies critically. Results are often quoted without determining if these results are valid and reliable. For example, Carver (1993:287) states that too many research results are blatantly described as significant, when they are in fact trivially small and unimportant. He says that there is no excuse for saying that a statistically significant result is significant because this language use erroneously suggests to many readers that the result is automatically large, important and substantial.

The result of this state of affairs is that many readers, when confronted with quantitative studies – especially those using statistics, either avoid reading the article, or take a short cut through it. Very often this entails skipping the 'Method of Research' section to get to the 'Discussion', where they try to find out what the study was all about and if they can find something useful to implement in their classrooms (Brown, 1991). By skipping the 'Method' section, however, readers not only miss the heart of the study, but also buy the author's argument without critical evaluation.

The purpose of this article is first to indicate some of the deficiencies in quantitative research studies very briefly, and then to provide guidelines for interpreting and evaluating the 'Research Method' section of quantitative studies within the field of second language research.

2. An analysis of quantitative research studies

In recent years considerable concern has arisen over the misapplication or avoidance of appropriate quantitative methods in language learning research. Brown (quoted in Henning, 1986), for example, expresses concern that established conventions in quantitative research methodology are not consistently adhered to by quantitative researchers in second language research.

Table 1 presents an analysis of the method sections of a number of quantitative studies investigating the influence of affective factors (such as personality, motivation, anxiety, competitiveness, etc.) on second language acquisition. This analysis highlights some of the weaknesses in these studies. It is clear that researchers, students and teachers need to be wary of quoting the results of studies at random without critically evaluating the studies.

Table 1: An analysis of the Method Sections of quantitative research on affective factors up to 1986

Researcher(s)	Subj.	Instr.	Var.	DCP	Des.	Anal.
Chastain (1975)	X	X	X		X	X
Dunkel (1947)	X	X	X	X	X	X
Gardner & Lambert (1959)	X	X	X	X	X	X
Gardner *et al.* (1977)	X	X			X	X
Gardner *et al.* (1979)	X	X			X	X
Kleinmann (1977)					X	X
Tucker *et al.* (1976)		X			X	X
Wittenborn *et al.* (1945)	X	X			X	
Young (1986)	X				X	X

Key:

Subj. = Subjects
Instr. = Instruments
Var. = Variables
DCP = Data Collection Procedure
Des = Design
Anal. = Analysis

The analysis in Table 1 is not intended to be detailed or comprehensive, but merely to illustrate that not all studies are perfect. The areas which are deficient are indicated by *X* in the table. It is obvious that most of these studies have a number of deficiencies. For example, the article by Gardner and Lambert (1959) reveals the following:

* **Subjects**

The researchers do not mention whether the subjects were selected randomly, whether they constituted an intact group or whether they were volunteers. Failure to address these issues will impede the generalizability of the results. The internal and external reliability and validity can therefore be influenced.

* **Instruments**

The researchers do not give any indication of the reliability and validity of the instruments used, nor do they mention whether all the tests were standardized. The researchers also made use of a few sub-scales of tests constituting a larger battery. However, very often the battery needs to function as a unit and by using only a few sub-scales the researchers may influence the reliability of the instruments.

* **Variables**

The variables used in the study are not clearly specified and operationalized.

* **Data collection procedure**

No information is given on how and when the researchers collected the data. No indication is given of the setting, the instructions given to the students or the time period needed for data collection.

* **Design**

The design is not specified; therefore it is impossible to determine whether the correct one was chosen.

* **Analysis**

The researchers do not mention whether the assumptions underlying the use of the statistical procedures employed were met. For example, correlations presume the use of normally distributed data.

The brief analysis presented above reveals that very often basic information which is essential for the methodology section is either not reported or buried away in the body of the article. It is interesting to note that a review of a few South African journals, which specifically focus on language learning, for example the SAALT (*Journal for Language Teaching*), reveals the limited number of studies concentrating on quantitative research. Most of the studies that have been conducted suffer from the same deficiencies mentioned in Table 1.

The rest of this article will discuss the components of quantitative research studies in second language acquisition and indicate how these studies can be evaluated and conducted. As stated above, the focus is on the Method of Research section. The purpose of this section is to explain how the study has been conducted. The standard rule is that the description should be thorough enough for a competent researcher to reproduce the study (Hatch & Lazaraton, 1991). The article will focus on the subjects, the instrumentation, the variables, the data collection procedure, the research design and the data analysis.

3. Subjects

This section describes how and why the subjects are selected and what characteristics they have that are pertinent to the study. Since language learning deals primarily with human beings, a large proportion of studies gather data about characteristics of designated human populations. The study itself will generally be directed to a particular population, but the researcher must decide which specific individuals (the sample) will provide the data. The key issue is *how* this group is selected. Are they randomly chosen from a larger population? Are they volunteers? Are there any special criteria used for choosing them? (Brown, 1988). The answers to these questions are important if one is to decide whether the results can be generalized to the field at large. The major criteria in evaluating these descriptions are precision and replicability (Hatch & Lazaraton, 1991).

Since the purpose of drawing a sample from the population is to obtain information concerning that population, it is extremely important that the individuals included in the sample constitute a representative cross section of the individuals in the population. That is, samples must be representative if one is to be able to generalize with confidence to the population. Various sampling techniques are available to the researcher: random sampling, stratified sampling, cluster sampling and systematic sampling. A problem that must be faced in planning every research study is to determine the size of the sample necessary to attain the objectives of the planned research. Technically, the size of the sample depends on the precision the researcher desires in estimating the population parameter at a particular confidence level. The best answer to the question of size is to use as large a sample as possible; a sample of 30 is usually accepted in experimental research (Ary *et al.*, 1972). According to Borg and Gall (1979) larger samples are necessary under the following conditions:

* When many uncontrolled variables are present.
* When small differences or relationships are anticipated.
* When groups must be broken into subgroups.
* When the population is highly heterogeneous on the variables being studied.
* When reliable measures of the dependent variable are not available.

4. Instruments/Materials

This section should give the reader a description of the instruments, materials, or tests used to collect the data. Teaching materials, questionnaires, rating scales and tests should be described in detail unless they are well known (Brown, 1988). Any other pertinent information, such as range of possible scores, scoring methods used, types of questions, and types of scales, should also be included.

Inasmuch as the instruments used will provide the operational definition of the variables, their use must be justified as being appropriate for that purpose. The researcher should explain why the instrument used was selected as the most appropriate definition of the variable under consideration. If an instrument is one already established, the researcher should include reported evidence of its reliability (consistency) and validity (what the test measures) for the purpose of the study. If the researcher is developing his/her own instruments he/she should outline the procedure to be followed in their development.

In the next section variables are briefly discussed. Variables do not usually form a separate heading in the Method section; however, a brief discussion of variables is included because they are so closely linked to the instruments section.

5. Variables

A variable is an attribute or set of observations that may vary, or differ in a study (Hatch & Farhady, 1982; Brown, 1992). Most research in the second language field is concerned with identifying the variables that are important to language learning and discovering how these variables affect the learning and teaching of languages. Five different types of variables can be distinguished according to the functions they perform in a study: dependent, independent, moderator, control and intervening variables.

A dependent variable is observed to determine what effect, if any, the other types of variables may have on it. In other words, it is the variable of focus – the central variable – on which other variables will act if there is any relationship (Brown, 1988).

Independent variables are variables selected by the researcher to determine their effect on or relationship with the dependent variable. An independent variable is one that is selected and systematically manipulated by the researcher to determine whether, or the degree to which, it has any effect on the dependent variable.

A moderator variable is a special type of independent variable that the researcher has chosen to include in order to determine if this moderator variable has an effect on the relationship between the independent and dependent variables.

It is virtually impossible to include all the potential variables in each study. As a result, the researcher must attempt to control, or neutralize, all other extraneous variables that are likely to have an effect on the relationship between the independent, dependent and moderator variables. Control variables, then, are those that the researcher has chosen to keep constant, neutralize, or otherwise eliminate so that they will not have an effect on the study.

The intervening variable may be used to describe the theoretical relationship between the independent and dependent variables. They are constructs that may explain the relationship between independent and dependent variables but are not directly observable themselves (Brown, 1992).

A number of problems can arise, both within and outside a study, that may create flaws in terms of the validity and reliability of the study, the degree to which a study and its results correctly lead to, or support, exactly what is claimed. The problems themselves result from extraneous variables that are relevant to a study but are not noticed or controlled. Brown (1988) discusses extraneous variables from four perspectives: environmental issues, grouping issues, people issues, and measurement issues. Table 2 gives a brief summary of some of the common problems experienced with extraneous variables.

Table 2: Extraneous variables: Potential problems

Focus	**Potential problems**
Environment	Natural variables (noise, temperature) Artificiality
Grouping	Self-selection Mortality (drop-out) Maturation (age)
People	Hawthorne effect Halo effect Subject expectancy Researcher expectancy

Measurement	Practice effect (taking same test twice) Reactivity (different pre- and posttest-standard) Instability of measures and results

(Adapted from Brown, 1988).

All variables must be operationally defined (Brown, 1988). An operational definition ascribes meaning to a construct by specifying the operations that must be performed in order to measure the concept. This type of definition is essential in research, since data must be collected in terms of observable events. An operational definition is very specific in meaning; its purpose is to delimit a term, to insure that everyone concerned understands the particular way in which a term is being used. It must be a definition that is based on observable, testable or quantifiable characteristics.

6. Data collection procedure

This section should describe how the data are obtained. All testing procedures for obtaining scores on the variables of interest should be explained. How tests are administered and who does so are important features. The setup of the testing situation and instructions given to the subjects should be noted. What were the environmental conditions like during the experiment? Were they the same for all the subjects involved? The answers to these and many other potential questions should make it possible for the reader to understand exactly how the study was conducted.

The 'procedures' section contains most of the detail that allows another researcher to replicate the study. Tuckman (1988) outlines these details, which generally include:

* the specific order in which steps were undertaken;
* the timing of the study (e.g., time for different procedures and time between different procedures);
* instructions given to subjects, and
* briefings, debriefings and safeguards.

According to Seliger and Shohamy (1989) it is important to use procedures which elicit quality data, since the quality of any research study depends largely on the quality of the data collected, and the quality is directly related to the data collection procedures.

7. Design

The research design refers to the conceptual framework within which the experiment is conducted. It is important to plan the research design because it will help the researcher determine how the data should be analysed. A research design has two very important functions: it provides opportunity for the comparisons required by the hypotheses of the experiment, and it enables the researcher through statistical analysis of data to make a meaningful interpretation of the results (Borg & Gall, 1979).

Design is the key to controlling the outcomes from experimental research. A well-designed study is one in which the only explanation for the change in the dependent variable is how the subjects were treated (independent variable). The design enables the researcher to eliminate all rival or alternate hypotheses. The basic types of research design can be divided into three categories: pre-experimental, true experimental and quasi-experimental (Campbell & Stanley, 1963; Borg & Gall, 1979). The type of design the researcher selects will depend on the hypothesis or research objective he/she has set for him-/herself. Each type of research design answers a different question. If the hypothesis the researcher is testing asks 'Does a change in the independent variable produce a change in the dependent variable?', then a true experimental design is required. However, a true experimental design cannot always be used, as variables are often difficult to control. One of the other research designs must then be used, but one should realize that this is not the 'ideal' design. Conclusions should only be drawn as data and research design permit. It must be borne in mind that research is limited because of the use of a research design other than a true experimental design (the 'ideal'). The limitations section of the article is the place to demonstrate that the researcher is aware of the fact that the research is not completely ideal. Table 3 gives an outline of some of the most commonly used research designs in the second language field.

Pre-experimental designs (cf. Table 3) control very few of the sources of invalidity. None of the designs has a random assignment of subjects to groups, they lack a control group, and they fail to provide for the equivalence of a control group (see table on the next page):

Table 3: Research designs

Pre-experimental	True experimental	Quasi-experimental
• One-shot case study	• Randomized posttest-only control-group design	• Nonrandomized control-group pretest-posttest design
• One-group pretest-posttest design	• Randomized matched subjects posttest only	• Time series design
• Static group comparison	• Randomized pretest-posttest control group design	• Counterbalanced designs
	• Solomon three-group design	
	• Solomon four-group design	
	• Factorial designs	

Notation

R – Random assignment of subjects to groups.

O – An observation or test (subscripts refer to the order of testing, that is, O_1 is the first time a test is given, while O_2 is the second test administration).

X – means a treatment is applied. (Subscripts X_1, X_2 on different lines refer to different treatments: subscripts on the same line mean the treatment is administered more than once; a blank space means the group is a control).

--- A dotted line between groups means the groups are used intact rather than being randomly formed.

- **One-shot case study: X O**

A group of subjects receive a treatment followed by a test to evaluate the treatment.

- **One-group pretest-posttest design:** $O_1 \quad X \quad O_2$

A group of subjects are given a pretest followed by a treatment period and then they are given a posttest to observe whether any change in performance has occurred.

- **Static-group comparison:** $\begin{matrix} X & O_1 \\ & O_2 \end{matrix}$

This design compares two groups, one of which receives the treatment and one of which does not.

In true experimental designs (cf. Table 3) the groups are randomly formed, allowing the assumption that they were equivalent at the beginning of the research.

- **Randomized posttest-only control-group design:** $\begin{matrix} R & X & O_1 \\ R & & O_2 \end{matrix}$

This design is similar to the static group comparison design except that the groups are randomly formed, therefore, allowing the conclusion that significant differences between O_1 and O_2 are due to X.

- **Randomized matched subjects posttest only**

This design is similar to the randomized posttest-only control-group design except that instead of using random selection to obtain equivalent groups, it uses a matching technique. Subjects are matched on one or more variables that can be measured, such as IQ, or placement test scores.

- **Randomized pretest-posttest control group design:** $\begin{matrix} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{matrix}$

In this design the groups are randomly formed, but both groups are given a pretest as well as a posttest. The major purpose of this type of design is to determine the amount of change produced by the treatment; that is, does the experimental group change more than the control group?

- **Solomon three-group design:** $\begin{matrix} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \\ R & & X & O_5 \end{matrix}$

This design is similar to the randomized pretest-posttest control group design, but it has the advantage that it employs a second control group and thereby overcomes the difficulty inherent in the randomized pretest-posttest control group design, namely the interactive effect of pretesting and the experimental manipulation. This second control group is not pretested but is exposed to the X treatment.

- **Solomon four-group design:**

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

This design provides still more rigorous control by extending the three-group to include one more control group. The purpose is explicitly to determine whether the pretest results in increased sensitivity of the subjects to the treatment. This design allows a replication of the treatment effect (is O_2>O_4) and (is O_5>O_6), an assessment of the amount of change due to the treatment (is O_2-O_1 > O_4-O_3), an evaluation of the testing effect (is O_4>O_6) and an assessment of whether the pretest interacts with the treatment (is O_2>O_5).

- **Factorial designs:**

R	X_1	O_1
R	X_2	O_2
R		O_3

A factorial design is one in which two or more variables are manipulated simultaneously in order to study the independent effect of each variable on the dependent variable as well as the effects due to interactions among the various variables. In this case, three levels of the independent variable exist, where one is the control and the X_1 and X_2 represent two levels of treatment.

The purpose of quasi-designs (cf. Table 3) is to fit the design to settings more like the real world while still controlling as many of the threats to internal validity as possible.

- **Nonrandomized control-group pretest-posttest design:**

O_1	X	O_2
O_3		O_4

This design is similar to the randomized control-group pretest-posttest design except that in this case intact groups are used (e.g., classes in school).

- **Time series:** O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

This design has only one group but attempts to show that the change that occurs when the treatment is interjected differs from the time when it is not.

- **Counterbalanced designs:**

	X_1	X_2	X_3	X_4
Group	A	B	C	D
	C	A	D	B
	B	D	A	C
	D	C	B	A

This design can also be used with intact groups, and it rotates the groups at intervals during the experimentation. All subjects receive all experimental treatments at some time during the experiment.

Control is the essence of the quantitative method. Without control it is impossible to evaluate unambiguously the effects of an independent variable. In order to be able to draw a conclusion concerning the relationship of the independent variable and the dependent variable, it is necessary to control the effects of any extraneous variables. An extraneous variable is a variable not related to the purpose of the study, but which may affect the dependent variable (Brown, 1988). *Control* is the term used to indicate a researcher's procedures for eliminating the differential effects of all variables extraneous to the purpose of the study. He controls, for instance, when he makes the groups comparable on extraneous variables that are related to the dependent variable. Other methods of control include: simple randomization, randomized matching, homogeneous selection and analysis of covariance. It is therefore important to take note of all the potentially influential variables (cf. section 5).

The design of the study determines what statistical techniques should be used, not vice versa. In other words, one decides what design will enable one to observe the hypothesized relationships, then one selects the statistical procedure that fits the questions asked and the nature of the data involved. The appropriate statistic to use is determined partly by the type of measurement scale characterizing the dependent variable.

8. Analysis

The data analysis procedure must also be reported. In most quantitative studies some type of statistical analysis is used. Typically, the researcher will explain the proposed application of the statistics. In nearly all cases, descriptive statistics are provided, such as means and standard deviations for each of the variables. If correlational techniques (relationships among variables) are used, then the variables to be correlated and the techniques are named. Statistical analyses have many variants, and choosing one variant over another can dramatically affect the results. So it should be clear to the reader exactly which analyses were used and in what order. In other words, the analyses should be explained just as they were planned, step by step (Brown, 1988).

Brown (1992) states that assumptions are preconditions that are necessary for accurate application of a particular statistical test. In some cases, these assumptions are not optional; they must be met for the statistical test to be meaningful. It should be clear to the reader that the assumptions were checked and met. A few of the principal assumptions discussed by Brown (1992) are the following:

The assumption of *independence of groups* implies that there must be no association between the groups in a study. The most obvious violations of this assumption occur when the same people appear in more than one group. A second assumption is *independence of observations*. This is often required for proper application of correlational and other statistics. Here, the assumption is that there is no association between the observations within a group. *Normality of the distributions* is often required for proper application of statistical tests in mean comparisons. Violations of this assumption are less troublesome if the sample sizes are large. The distribution can be taken as normal if there is room for two or three standard deviations on either side of the mean and if there are no outliers (extremely large or small values). Violations of the assumption of *equal variances* can be detected by examining the standard deviations in a study because the variances are simply the standard deviations squared. If there are big differences in these squared values, there are probably violations of this assumption. The assumption of *linearity* often applies in the correlational and prediction family of statistics. It means that there is a straight-line relationship between the two variables involved. This assumption can be checked by examining a scatterplot of the two variables. The assumption of *nonmulticollinearity* is a problem if the variables in a study are too highly interrelated. This assumption can be checked by examining a table of the correlation coefficient for each pair of variables in the study. The final assumption of concern is that of *homoscedasticity*. This assumption, which is often applied to statistical procedures based on correlation and prediction, is that the variability of scores on one variable is about the same at all values of the other variable. This assumption can also be checked by examining a scatterplot of the variables involved.

9. A checklist for the evaluation of quantitative studies

Table 4 contains a checklist that researchers, students and teachers may find useful when writing or reading and evaluating a quantitative research study.

Table 4: Checklist for the evaluation of the method section of a quantitative research study

Method of research	Checklist
1. Subjects	Is the description of participants adequate? Is the method of selection clear?
2. Instruments/Materials	Is there a description of tests, questionnaires, rating scales, etc.? Is the reliability and validity of tests used, indicated?
3. Variables	What types of variables were involved? Which were the dependent and independent variables? Were they properly labelled? Are there any extraneous variables that were not accounted for in the study? Has the researcher successfully controlled them or were they unnoticed?
4. Data collection procedure	Is there a description of the preparation of materials, scoring, administration, and so on? Is there a description of the conditions during the study? Is the timing of the procedures specified?
5. Design	Was the design used by the researcher specified? Was the proper design used to test the hypothesis under investigation?
6. Data analysis	Is there a description of the arrangement and grouping of the data? Are the statistical tests listed in order of use? Did the researcher indicate whether the assumptions underlying the statistical tests that were used were met? Are the statistics appropriate for the design?

10. Conclusion

Reading and interpreting quantitative research in the second language acquisition field is, or should be, a creative and critical exercise. According to Brown (1988), it is creative in the sense that the reader must actively participate with the original researcher. It is, therefore, important that the study must be replicable. This is perhaps the single most important yardstick to hold up against any study.

Teachers and students must be able to evaluate studies critically. If, for example, the basic research design or the primary statistical tests are faulty, the results may be meaningless. A sophisticated and critical audience can only help to improve research in the second language field.

Brown (1991) states that there are no guarantees that the articles that appear in print are 100% correct or uncontroversial. It is therefore the student and teacher's responsibility to read any articles that interest them as carefully and critically as they can so the the interface between teaching and research can be strengthened.

Bibliography

Ary, D., Jacobs, L.C. & Razavieh, A. 1972. *Introduction to Research in Education*. New York : Holt, Rinehart and Winston.

Borg, W.R. & Gall, M.D. 1979. *Educational Research: An Introduction*. (3rd Ed.). New York : Longman.

Brown, J.D. 1988. *Understanding Research in Second Language Learning*. Cambridge : Cambridge University Press.

Brown, J.D. 1991. Statistics as a Foreign Language – Part 1: What to Look for in Reading Statistical Language Studies. *TESOL Quarterly*, 25(4):569-586.

Brown, J.D. 1992. Statistics as a Foreign Language – Part 2: More Things to Consider in Reading Statistical Language Studies. *TESOL Quarterly*, 26(4):629-664.

Campbell, D.T. & Stanley, J.C. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago : Rand McNally.

Carver, R.P. 1993. The Case against Statistical Significance Testing, Revisited. *Journal of Experimental Education*, 61(4):287-292.

Chastain, K. 1975. Affective and Ability Factors in Second Language Acquisition. *Language Learning*, 25(1):153-161.

Dunkel, H.B. 1947. The Effect of Personality on Language Achievement. *The Journal of Educational Psychology*, 38:177-182.

Gardner, R.C. & Lambert, W.E. 1959. Motivational Variables in Second Language Acquisition. *The Canadian Journal of Psychology*, 13(4):266-272.

Gardner, R.C., Smythe, P.C. & Brunet, G.R. 1977. Intensive Second Language Study: Effects of Attitudes, Motivation and French Achievement. *Language Learning*, 27(2): 243-261.

Gardner, R.C., Smythe, P.C. & Clement, R. 1979. Intensive Second Language Study in a Bicultural Milieu: An Investigation of Attitudes, Motivation and Language Proficiency. *Language Learning*, 29(2):305-320.

Hatch, E. & Lazaraton, A. 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. New York : Newbury House.

Hatch, E. & Farhady, H. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, MA. : Newbury House.

Henning, G. 1986. Quantitative Methods in Language Acquisition Research. *TESOL Quarterly*, 20(4):701-708.

Kleinmann, H.H. 1977. Avoidance Behavior in Adult Second Language Acquisition. *Language Learning*, 27(1):93-107.

Seliger, H.W. & Shohamy, E. 1989. *Second Language Research Methods*. Oxford : Oxford University Press.

Tucker, G.R., Hamayan, E. & Genesee, F.H. 1976. Affective, Cognitive and Social Factors in Second Language Acquisition. *The Canadian Modern Language Review*, 32:214-226.

Tuckman, B.W. 1988. *Conducting Educational Research*. (3rd Ed.). New York : Harcourt Brace Jovanovich.

Wittenborn, J.R., Larsen, R.P. & Mogil, R.L. 1945. An Empirical Evaluation of Study Habits for College Courses in French and Spanish. *The Journal of Educational Psychology*, 36(8):449-474.

Wolfson, N. 1986. Research Methodology and the Question of Validity. *TESOL Quarterly*, 20:689-699.

Young, D.J. 1986. The Relationship Between Anxiety and Foreign Language Oral Proficiency Ratings. *Foreign Language Annals*. 19(5):439-445.