Revista Alicantina de Estudios Ingleses

nº 17, November 2004

Departamento de Filología Inglesa
Universidad de Alicante

ISSN 0214-4808 ● CODEN RAEIEX

The *Revista Alicantina de Estudios Ingleses* is published yearly by the Department of English at the University of Alicante in volumes of approximately 250 pages. The journal aims to provide a forum for debate and an outlet for research involving all aspects of English Studies.

NATURE AND FORMAT OF THE ARTICLES: The *Revista* would welcome articles of the following kinds: (1) Articles on linguistics and ELT, literature, literary theory and criticism, history and other aspects of the culture of the English-speaking nations. Articles should not exceed nine thousand words in length. (2) Bibliographies of studies on *very* specific topics, providing a brief introduction and a list of basic publications. A concise index of contents may optionally be included. (3) Reviews and review articles on recently published books in the field of English Studies. (4) Poetry translations (English-Spanish and Spanish-English). All articles submitted should follow the guidelines which can be obtained from the following Internet address:

http://www.ua.es/dfing/publicaciones/raei/general/instrucciones.htm

Manuscripts should include an abstract in English of about one hundred words in length. In normal circumstances, the editors will only consider for publication those contributions written in English and recorded on disk. Two print-outs of the contribution should also be included. Articles will only be returned at the authors' express wish, if so requested at the time of submission. All correspondence should be addressed to:

Revista Alicantina de Estudios Ingleses, Departamento de Filología Inglesa, Universidad de Alicante, P. O. Box 99, E-03080 ALICANTE (Spain)

● ADVERTISING: The journal will be pleased to carry advertisements in either full-page (17 x 24 cms. approx.) or half-page (17 x 12 cms. approx.) format. Prices and information are available on request at the above address

● EXCHANGES: The *Revista Alicantina de Estudios Ingleses* will be happy to make exchange arrangements with similar journals in the same field. All such proposals should be made in writing to the above address

● SUBSCRIPTIONS: The price of subscriptions for FOUR issues of the *Revista Alicantina de Estudios Ingleses* is as follows: (1) in Spain, 25€ for libraries and institutions, and 18€ for private subscribers; (2) in countries other than Spain, US $30 for libraries and institutions, and US $25 for private subscribers.

Correspondence on subscriptions should be adressed to:

Marcial Pons Librero
San Sotero 6
28037 MADRID (Spain)
revistas@marcialpons.es
Tel.: +34 913043303
Fax: +34 913272367

Portada: Enrique Pérez
Gabinete de Diseño de la Universidad de Alicante

ISSN: 0214-4808

Depósito Legal: A-22-1989

Edición de: COMPOBELL, S.L. Murcia

Estos créditos pertenecen a la edición impresa de la obra

Edición electrónica:

Espagrafic

A Research Program for Foreign Language Teachers: Building up Investigation in Classroom Contexts

G. Luque-Agulló and M.M. Ramos-Álvarez

Contents

A Research Program for Foreign Language Teachers: Building up Investigation in Classroom Contexts

G. Luque-Agulló and M.M. Ramos-Álvarez
University of Jaen
gluque@ujaen.es

Abstract

When researchers in the field of Applied Linguistics or L2 teachers try to carry out any kind of research, several problems arise. The first one is knowing how to start the process of 'researching'. The second problem, perhaps more frustrating, arises when one finds out that one cannot grasp the complexities to be taken into account when getting started on the research task at hand. Third, one has to find out how to adapt the different steps of the research to one's own situation, context, and degree of knowledge, so that data can later be able appropriately coded and analysed in order to establish some sort of conclusion. Last but not least, a further struggle tends to appear on discovering that what should have been done and what has actually been done are two very different things. This paper aims at improving this situation by establishing in a clear-cut way the three levels any investigation should include –conceptual, methodological and empirical–. Then each framework –level– is described, producing a very specific and up-to-date research agenda in which every specific notion is defined and justified.

1. Introduction

When teachers devoted to the instruction of second (or foreign) languages want to carry out any kind of research, either classroom-oriented or within the classroom (Note: distinction taken from Nunan, 1992), several problems arise. The first problem is the question of knowing how to start and go about the process of 'researching', as most L2 teachers usually have very sketchy notions of how to investigate (which is not their fault: knowing the second language, about the second language, and how to teach it are wide enough fields of knowledge).The second problem, no doubt more frustrating, arises when getting started on the research task at hand and finding out that the complexities of the task cannot be managed. Third, once these teachers, probably already in anguish, have grasped some of the intricacies of this task, they have to discover how to adapt the different steps of the research to their own situation, context, and degree of knowledge, so that data can later be appropriately coded and analysed in order to establish some sort of conclusion. Last but not least, a further struggle tends to appear when discovering that what one should have done and what one has actually done are two very distinct and different things. **(note 1)**

This work tries to help overcome some of these problems, by making a brief summary of the steps involved in research tasks and projects (there is a lot of material related to this topic, see, among others: Nunan, 1991, 1992; Seliher & Sohamy, 1989; Van Lier, 1988; Manly, 1992; Maxwell and Delaney, 1990), the possible options and choices included in every step of this process, and some of the pitfalls that can be avoided along the way. Thus we do not find ourselves in the situation described in one of our endnotes, nor does one have to bear the criticisms of *pure* researchers or 'scientists' who, from the standpoint of their ivory towers (see a very good comment in Pedhazur and Schmelkin, 1991, on the ironies derived from distinguishing pure and applied research), would probably disregard the investigation carried out by teachers or applied linguists and consider it as biased and in need of confirmation by means of *pure scientific research and real experimental methods*.

Our aim here is the identification and explanation of the steps involved in implementing research studies, putting a special emphasis on the integration of *all* the components through an up-to-date approach. The process of scientific research has been described as a set of *hierarchised* operations or stages that can be grouped in three different levels: the theoretical-conceptual, the technical-methodological and the statistical-analytical (see Arnau, 1989).

1. Conceptual framework

In order to do this, we will briefly outline a general model for doing research, listing the courses of action involved. This framework could be adapted to related fields, although we will use examples and applications for classroom research, (including both classroom research and classroom oriented research).

When a teacher or applied researcher wants to do some research, there usually exists the underlying motivation of finding out the answer to a **problem, idea or difficulty** that has arisen in connection with one's teaching, experience as learner, reading, or all three. The so-called problem triggers the need to delve into the complexities of unexplored realities; in other words, those phenomena for which the teacher/researcher and current research do not have an answer. This problem has an unfocused formulation, that is to say, it has the form of a somewhat vague idea expressed as a proposal of an imprecise linguistic nature, so that it would have to be refined and rephrased in order to ultimately be considered as a research hypothesis. This problem has not only arisen from the researcher's mind, but evolves from a particular context and socio-economic situation, within a theoretical framework that is valid at the moment the problem appears, takes into account a particular range of research strategies and entails the acceptance of a given empirical domain (a set of data obtained by previous research **(note 2)**). It is also determined by the readily available technology at the disposal of the *investigator*.

For example, a problem or general question may be to wonder *how deaf students can learn English as a foreign language*. This problem may have appeared because the teacher/researcher has one deaf student in his class. As we can see, the socio-economic situation and context within which the researcher is located determines the kind of problems that are the issue of his enquiry. He will have read about deaf people, so that when he starts to work on the problem he is accepting a set of already-proven data, and all these data will be connected with a given theoretical framework that depends on the historical point in time in which he is living. Researching in the fifties (with behaviourist positions) would not imply the same theoretical paradigm as researching in the nineties (with cognitive or connectionist accounts). In connection with the problem at hand, the teacher/researcher will also have taken into account (consciously or unconsciously) the technology accessible within his means, situation and background, so that his choice of research strategies will depend on aspects such as the possibility of being able to use a laboratory, listening aids and the like.

Once the problem has been established, general ideas and concepts have to be specified in a clear-cut way. That is to say, the teacher/researcher has to proceed from a given **Theory or Model** to his own **Hypothesis,** the latter being concrete and precise enough to allow later transformation into a **prediction**, which would have to fulfil two conditions: it has to allow being contrasted with the data and it can be *potentially* replicated. **(note 3)** For the previous example, *the acquisition of L2 by deaf students,* the teacher would have to place himself/herself in a given theoretical framework and then try to reformulate his/her problem so that it turns into an *operative* hypothesis. The Model within which the teacher/research-er places himself/herself is a way of understanding the world at large, some of its phenomena and the research strategies chosen to study those phenomena. The theory also comes with a set of general **laws** (note 4) that allow the researcher to **deduce** concrete and operative **hypotheses**. In addition, the theory or model includes a set of hypotheses and predictions about the phenomena it attempts to explain. For instance, if a teacher/researcher paced himself/herself within the UG (Universal Grammar) model, he/she would be accepting a very definite way of looking at linguistic phenomena, discarding other forms of linguistic analysis. Within this framework the teacher would also assume the truth-value of some **assumptions or constrictions**, such as the existence of a Language Acquisition Device (LAD) or the universality of some linguistic features. These assumptions are accepted ideas or attitudes within the theory which are not questioned by the scientific community, and in a metaphorical sense they *shield* the core formed by the **laws and enunciates** of the model (i.e. within mentalist models, language is assumed to be an innate quality of human beings: there is a language Acquisition Device –LAD–). If these assumptions were to be chal-lenged by other theories or scientists, the theory would be liable to be contested and replaced by other alternative models. To continue with our example, if the existence of the

LAD began to be questioned, the whole UG theory would be at stake and would probably be discarded. **(note 5)**

Up to now the teacher/researcher is supposed to have travelled from the more general, abstract and theoretical framework to a more concrete one, though this does not show the whole picture. In fact, a process of going back and forth between three levels should guide every research question: the theoretical, the methodological and the empirical ones, the latter including observable phenomena or data. In other words, the theory within which we place ourselves restricts the data, as the researcher will observe only those phenomena that can be accounted for by his chosen model. Conversely, data can sometimes limit or even modify the theories so that they have to be completely reformulated, or, then again, the appearance of new data might compel researchers to revise the methodology of research to be used within a given model. As an example related to the interaction between the presence of new data and the subsequent change in methodology, recent research methodologies on neurolinguistics, such as Regional Cerebral Blood Flow (RCBF) have promoted the finding of new evidence on lateralisation processes and the location of different linguistic tasks in the brain, in turn confirming hypotheses within this theory and rejecting others. In the field of second language acquisition, detailed classroom studies with complete transcriptions have frequently compelled researchers to use new methods of analysis, from grammar to discourse-based approaches, or, very occasionally, the presence of new evidence has stimulated the reformulation of accepted hypotheses. Looking at data –children learning to speak the second language– Hatch (1978) found out that second language learners started to acquire the language by speaking, and from that process of communication grammar patterns arose, instead of the other way round, which was what researchers (and, by implication, teachers) believed up to that point in time. In general, finding new evidence does not always lead to a change in the conceptual framework, as it is a rather more complex process, in the sense that the changes in any of the three levels may subsequently produce a revision of the other two.

Once the teacher/researcher has reached the juncture at which he/she is able to draw hypotheses, a further predicament arises: *What is a* ***Hypothesis***? Let us begin by saying what cannot be considered as a hypothesis: it is not a general assertion; it has to be capable of being tested. Also, it has to follow from a theory, but with a clear empirical orientation. This fact implies the need to be accurate in the linguistic formulation of hypotheses. To illustrate this point, one cannot formulate a hypothesis by saying (A) *Input promotes acquisition in a second language* **(note 6)** as it cannot be proven. Rather, we would have to reach a higher degree of accuracy by following the form –whenever possible– of a causal relationship (If…then…), either explicitly formulated or with an implicit *If-then* relation-ship. Thus, one could reformulate this hypothesis by saying (B) *if we simplify input with elaboration and visual aids, then students will understand the narrative discourse better*. However, the hypothesis is yet not very concrete, as we would have to specify what elaboration is, what kind of visual techniques we have in mind, and show how com-prehension can be measured so that it really shows an improvement. A more concrete hy- pothesis would be (C) *for students at pre-syntactic levels, the frequency of use of formulae within the narrative discourse is higher than that of grammatical markers*.

In this case, we can see the word *frequency*. There is, then, a statistical or parametric prediction now. One can see the research consequences that this hypothesis would imply. That is to say, by having a good measuring tool for formulae and another for grammatical markers, the teacher/researcher could start by finding out the relationship between the two.

Hence, we have gone from a **research hypothesis** towards an **operative** one, which establishes the need to initiate the second stage in the research: the methodological framework.

2. Methodological framework

This section tries to show the diverse research strategies and basic aspects which the implementation of a particular study entails. A general model of behaviour is proposed and exemplified in order to guide choices in research.

2.1. Research strategies

There are two general groups of strategies, according to the degree of control they imply, including a very relevant intermediate option for our interests. The three-fold distinction between highly controlled or experimental situations, quasi-experimental, and non-controlled situations will be maintained throughout this section.

Once we have an operative hypothesis that can be put to the test from an empirical point of view, the teacher/researcher will have to make another set of decisions concerning the research agenda he/she should use. Specifically, the context where he/she is situated, the time available and other socio-economical determinants will determine the type of research to carry out, obviously within a theoretical (note 7) framework. The first decision that may determine the type of research strategies and research design chosen concerns the situation, which can be **controlled** or **not controlled**. Controlled situations are connected with experiments in the traditional sense, including direct manipulation and an exhaustive control of all the variables. They involve direct and complete monitoring of all the steps in the process, choosing the place, time, number of subjects randomly selected, and so on. Within the educational context, the handling of this kind of situation and subsequent analysis would entail using control techniques for known but unexpected variables that may arise, and a criterion of **random selection** for counterbalancing the effect of unknown variables. For example, if a teacher wanted to find out the effect of a teaching method, he would have to control, by randomly selecting and distributing the participants of the study, the known but unexpected variable of having some of his students attending private English lessons, which is very fashionable nowadays. With this distribution the teacher might also forestall the effect of other variables of this type, such as the level of intelligence of participants, and even prevent the possible effect of unknown variables. Provided this was achieved, the situation could be said to be highly controlled. But random distributions involve discarding a given class grouping (two classes) and replacing it with randomly created control and experimental groups. To be more concrete, let us mention a teacher who wants to know about the effect of the strategies students use within different teaching approaches (e.g. the task-based approach and one with a focus on form). If the teacher decides to use a controlled situation, he/she will have to reorganize his students in two groups at random, which means the groups will not coincide with the two class groups, consequently discarding the possibility of doing the research while he is actually teaching. This in turn entails the need to do research *outside* his and the volunteers' working hours, which is a lot to ask of students and teachers/researchers.

And still, some more unexpected or unknown variables might arise, and the ability of the researcher to foresee and prevent their effect determines the degree of control exerted on the situation. Usually, in the educational context it is very difficult to be in command of all the variables, and that is the reason why most research carried out in educational contexts tends to be quasi-experimental (with a high, but not complete, degree of control). To continue with our example, if the teacher decides to use one classroom for the task-based method and another for the focus-on-form one, the degree of control would not be very high, as he/she could not prevent the possibility of having one classroom with lots of very intelligent students, and another classroom without them, or one classroom with a high incidence of attendance private lessons, among other unforeseeable variables.

Once the choice is made and the teacher finds himself/herself within the framework of a quasi-experimental design, there are also a number of choices. But before examining those choices, it is important to mention that this type of research can be as valid from a scientific perspective as that carried out in controlled situations, particularly if the teacher/researcher has a clear notion of his aims and situation, recognises his limitations, and places himself/herself in the position of following all the appropriate steps of the design chosen. We will focus on the not-so-extremely controlled situations because the teacher is likely to find himself/herself in this circumstance for several reasons: first, classrooms are difficult areas of experimentation, as many unknown variables may arise. Second, it is very difficult to create a randomly selected control and experimental group, especially when this entails extra working hours for teacher and participants. Third, field experiments and quasi-experiments (both, strictly speaking, not controlled situations) may have more external validity **(note 8)** for the situation at hand, since, after all, the teacher/researcher may want to apply the answer to his research question to a particular context: his/hers, and other classes. In addition, this type of research can shed light on a number of phenomena inaccessible for more controlled situations, either because it sometimes may not be feasible, ethical or cost-effective, or, more important, because the participants' behaviour will probably be different in a natural and in an artificial context (e.g. a laboratory may inhibit the participants' spontaneous oral expression).

Within a high-but-not-completely controlled situation, we find **field experiments** (i.e. to research spontaneous speech in the L2 we would observe the interactions with classmates outside the class) and **quasi-experiments** (e.g. we set up two groups within the class). If the situation has a lower degree of control, we would be talking about qualitative strategies, either one that has happened through the **direct intervention** of the researcher or one in which **no intervention** has been effected**.** When one of these situations has a not-so-high degree of control, it may entail the presence of indirect manipulations or assignation/attribution variables, which are those unavoidable variables not determined by the researcher, but whose effect can be circumvented by deliberate distributions, so that we can investigate their influence. Besides, there might be other known or extraneous variables which are relevant to our situation –as seen in previous research– and that can be explicitly controlled. Thus, a quasi-experiment is characterized by indirect variables and a high degree of explicit control on extraneous variables. For instance, in a classroom situation, we might want to research the effect of attending private L2 lessons, which entails an indirect manipulation, as one cannot arbitrarily modify it. We might also consider that the students' intelligence could affect –or confound– the effect of that attendance, so that we should control this variable. As private classes have already given groupings, random distribution is not possible, thus making an experimental study impractical.

Less controlled situations also involve a number of choices. First, we may, as in experimental studies, want to see the effect of some manipulation on a given –already defined– situation. This is termed **qualitative study with intervention**, as the researcher intervenes with an action to bring about a result predicted in a set of hypotheses. For example, if a teacher wanted to find out about the effects on the students' proficiency of two teaching methods, he/she would have to intervene by devoting some time to teaching each method to a given group or looking for someone who did so (intervention, in both cases), after having checked the students' proficiency before the process started. This would take time and several measurements, stopping to see if the intervention is having or has had a result or effect. This research strategy is called, in specific terminology, interrupted temporal/time series. That is to say, the teacher would, after an initial test, apply each different method throughout a number of weeks or months, interrupting his/her intervention every few days, weeks or months to check on the effects of his/her 'manipulation'. After a given time and several measurements, he/she would have to decide when to end his/her intervention and then start to see if his initial hypotheses were fulfilled by analysing the data obtained.

Less controlled situations also allow for another course of action: **qualitative study without intervention**. This involves an investigation in which the researcher does not manipulate any variable, so that no apparent change is produced on his part to bring about a result on the participants. This situation implies looking at data from a descriptive point of view, by observing, making correlations, interviews, surveys, gathering and analysing general documents or carrying out longitudinal studies, always bearing in mind that no intervention on his/her part is expected or even allowed. The implication that follows is that, by using an interpretative tool (created by the teacher or obtained through current research literature), the teacher/researcher proceeds to study the behaviour of a group of people for a given phenomena in a given situation. Sometimes this type of study does not use any kind of interpretative tool, in the sense that the researcher only observes, takes notes, and so on, although there should always exist some process of *interpretation* in order to *organize* the observed reality or evidence into a meaningful set of conclusions when the last stage of the research is carried out. As an example of non-interventionist studies, we could mention the research devoted to how children acquire their mother tongue. In these studies there was no intervention, and the researcher just devoted a long time to recording, transcribing and analysing the samples taken when little children interacted with their families, peers or teachers.

A second type of study within the non-interventional type involves those studies that measure an event or phenomena for all the population. For instance, a survey company may want to find out what prospects the current president in a nation has for the next elections, or see the effect a serious event has had on the current view of a given population and their subsequent actions. This type of study will not be further contemplated within this paper for obvious reasons.

2.2. The Design of the research study

Once the teacher/researcher has decided (or his context has decided for him/her) on the type of research strategies he/she will be using, he/she must actually make a concrete *design* that includes, within the adscription to a given theory, **a plan, a structure and a strategy**.

The **plan** refers to how one may anticipate general questions regarding the anticipation of expected outcomes arising from the predictions, and the type of statistical analysis involved. It also includes the general shape of the experiment: what will be manipulated, measured and how relevant variables will be controlled.

The **structure,** more concrete**,** refers to the independent and dependent variables to be measured and manipulated, and the subsequent data collection technique **(note 9)** to be used. Independent variables (IV) are the ones that are manipulated, whereas dependent variables (DV) are those measures that *change as a consequence of* the former ones. For example, the dependent variable of '*knowledge of passive voice in a second language measured through an objective test*' may change as a result of the independent variable of '*type of explicit teaching*'. Thus, we could predict that '*focus-on-form'*, as a type of explicit teaching –IV–, might improve the '*knowledge of passive voice'*–DV– more than another type of explicit teaching (i.e. focus-on-forms). We should distinguish the levels *as values or variations* of the independent variable from its definition. In general, independent variable levels will be termed 'conditions'.

Third, the design must include the **strategy** to follow in order to solve the original problem**.** It implies carrying out the designed tasks, procedures, data collection instruments to measure the evidence, and control techniques to neutralize the effect of extraneous variables on the dependent variables. The tools the teacher/researcher will use to gather data depend on the type of strategy chosen. When the investigation has a lower degree of control, several simultaneous collection procedures should be used and combined. Classroom data gathering procedures include: think aloud techniques, diaries, open and closed interviews, observation, reports, questionnaires, surveys, and elicitation tasks. In short, how the dependent variable will be measured. Control techniques include, among other procedures, randomisation, blocking and counterbalancing. The first of these is used for strict experiments and the other two are used for quasi-experiments, where random selection is not possible.

In the third section we will deal with the last level of our research agenda.

3. The Statistical-analytical framework

Once the study has been carried out, there will be a corpus of evidence (the data from observable phenomena) to analyse and *interpret*. **(note 10)** Data obtained through a research study must then be analysed using statistical techniques, so that we arrive at an adequate interpretation. Normally, initial raw data can be massive, appearing scattered and sometimes affected by measurement errors. However, with the appropriate statistical analysis, we can select those aspects in data which are relevant in connection with the hypotheses or predictions we are trying to test. The analysis allows us to conclude if data are coherent with our starting hypothesis, or, conversely, if they are inconsistent. Conflictive evidence or data might have made the researcher alter the method (design) he/she intended to use in previous steps, or, exceptionally, readjust the theory which constituted his/her starting framework.

The corpus of evidence obtained will depend on the type of study carried out. If it is a descriptive one, some variable will have been measured. For example, if the teacher wants to measure L2 production in his/her students' output, this type of study indicates that he/she will be only interested in *describing* that performance, and in order to do that the teacher/researcher will have to be very careful creating a register to measure that performance, because that is the only way in which data will be evaluated. If the study belongs to the experimental framework, then measured variables are considered **depend-ent** ones (DV), as they will have changed as a result of the manipulation of some independent variable. For instance, the teacher/researcher may think telling stories (IV) will increase the students' performance (DV), so he/she would have to manipulate the teaching method, dividing students into two groups: one of them listens to stories with some support and another group which receives no support. Then he/she measures their performance (DV) and compares it. The third option considers quasi-experimental or correlational (nowadays called covariational) studies, which include **criterion variables**, similar to dependent variables, and **predicting variables**, similar to independent ones. For instance, a teacher/researcher may think there is a relationship between *age* and *L2 performance*, and he/she measures the two types of variables. The difference with a descriptive study is that *age* as a variable is believed to predict the second variable: *L2 performance*. Age would be a predicting variable whereas L2 performance would be the criterion one. The difference between this kind of study and the experimental one is that there is no manipulation of the variable (age), as it has only been measured to see its relationship with the second one (L2 performance). The data obtained through the use of a particular type of study must then be analysed.

Traditionally, the type of research strategy has determined the selection of different statistical techniques to analyse data, but this might be misleading, as it is the type of research study which determines how to interpret the *output* –or results– of those statistical techniques, rather than the other way round. If data arise from an experiment, causal conclusions are to be drawn, but this is not the case when the other research strategies are used. In fact no single technique exists that could be associated to each type of study, although we may find, among others, several options, including EDA, t-test, ANOVA/ ANCOVA/MANOVA, Regression, Categorical and log-linear, Cluster, Factor, Discriminant, Time Series, Structural Equation and Path Analysis. We have mentioned some frequently used statistical techniques, and in the following pages we will try to exemplify their use within a common statistical framework.

Nowadays there is a tendency to combine different statistical techniques within a single framework: the General Linear Model –GLM– and the Generalized Linear Model –GLZ– (Estes, 1991; Harris, 1994; Judd, and McClelland, 1989; Judd, McClelland, and Culhane, 1995), in which the implementation of the analysis arises from the nuclear principles of multivariate linear regression (linear principles when several dependent measures, or a multivariate context, might exist simultaneously). That is to say, the analysis is developed on the basis of a very easy structural equation, where data are equivalent to Model plus error. This Model we are using for analysing the evidence includes a statistical operativisation of the conceptual-hypothetical level, in such a way that the hypotheses and conceptual framework that initiated our research are *turned into* an *equation*. **(note 11)** The *error* within this perspective is equivalent to that part of the data that our model-working hypothesis is not able to explain. For example, if we have manipulated to teaching methods, the statistical model establishes that the L2 mean outcome (the students' production after the manipulation) will be different with each method. The philosophy underlying this approach is clear-cut and *clean*, as it means that, over time, only the model that best explains the data will be available. In other words, the most appropriate among several models is the one which best explains the data or has the lowest error rate.

The different levels of analysis which are used to *understand* the data must be consi-dered below. Note that we mention *levels* and not *types* of analysis, as the evidence should be explained by means of the different levels, which complement each other in the process of arriving at actual *proof*.

3.1. Descriptive level of analysis

The initial part of any design, whether experimental or qualitative-descriptive, consists of the description of data. The *massive* amount of information obtained has to be condensed, summarised and synthesized. To achieve this, we can use statistical measures which describe fundamental properties in the data, such as the mean, the Standard Deviation-Variance or The Standard Error of Mean –SEM–; (see Loftus and Masson, 1994). That is to say, we will be using properties such as the mean execution and variability rather than using the raw data we measured at the beginning of our study, to make them more manageable and comprehensible.

When data adopt an asymmetrical form, or they show outliers and extremes, it is better to use *robust* measuring alternatives, such as the Median, the Geometric Mean or the Harmonic Mean. For instance, when trying to measure reaction times in responding to a prime word in English, there is a tendency towards very low or very high reaction times occurring at a very infrequent rate, thus the asymmetry. Statistics such as the median are less sensitive towards those 'peaks' than classic statistics such as the mean (for a revision on reaction times in psycholinguistics, see Juffs, 2001).

As we have formerly mentioned, the descriptive level of analysis implies describing and synthesizing the shape/form data adopt, and this can be carried out through a graphical analysis termed Exploratory Data Analysis –EDA–, as appropriate graphics can highlight the shape adopted by data (see Behrens, 1997 for a review).

Finally, and within the descriptive framework, it is usually necessary to go further than obtaining means, deviations or graphs. To continue the synthesis and organization of the data we can use Data Reduction Methods such as Factor Analysis or Cluster Analysis. The first type can be considered when our aim is to extract some abstract dimensions or factors from the original data. For example, when we build a questionnaire to measure something, then the items in this questionnaire would be grouped around a number of factors. The second type –Cluster Analysis– explores how data tend to group, or in other words, it involves organizing observed data into meaningful structures, or developing taxonomies. The cluster can either be of the variables of the study or of the cases (individuals/participants) which have been measured. One of the best examples of Cluster Analysis in our field is the study by Biber (1988), which tried to see the features of different linguistic genres. By analysing large corpus of texts, and taking into account several linguistic dimensions, several categories (clusters) were obtained, or, using his terminology, text types, following the graphical form in which the evidence tended to be grouped.

To come back to the beginning and our General Linear Model (GLM), though the techniques illustrated above are sometimes used for exploratory purposes (i.e. to obtain empirical taxonomies), it is also usual to apply them with a confirmatory aim. That is to say, we may have a hypothesis which has initially pointed towards the number of factors or clusters in the data, and in accordance with this hypothesis, we then look to the unexplained *variance percentage* (i.e. called eigenvalues in the Factor Analysis technique) which remains unexplained after re-grouping the data according to the number of chunks of our hypothesis. For instance, if our questionnaire has to be contrasted with other questionnaires of the same type, then the most appropriate one will be that which achieves the larger reduction of the error following from the number of factors hypothesised (or Major Relative Explained Variance Percentage) .

3.2. Inferential Level of Analysis

Beyond the descriptive level, the final aim of most scientific research is to arrive at causal explanations, that is to say, researchers want to find out whether the dependent variable (DV) measured is affected by the manipulation of the independent variable (IV) according to our starting hypotheses. Or, to cut a long story short, most of us want to see the truth value behind our untested beliefs: to *really* know whether *doing something* to somebody –IV– has a direct effect –DV– on the behaviour of a person or thing. Taking again an example from our context, we may want to find out if one particular teaching method has a better result on the students' improvement of their L2 than other ones. We would choose two classes, one in which we apply the innovative method (the independent variable –IV– is the type of method, with two possible values to compare) and the other one which acts as the control group, with a traditional method. Then we would measure the students' proficiency level (DV) in the two classes and compare the results, which should be better for the group which was taught with the innovative teaching method, following the original hypothesis.

The problem of statistical inferences is that obtained results should allow generalizations. The resources involved in carrying out a research study entail an enormous effort which would only be worthwhile if conclusions can be generalized to other populations besides our participants (i.e. our teaching method proves to be useful for other students besides our own).

There are some scientific disciplines, such as physics, that do not require inferential developments, as their studies allow for highly controlled situations, error magnitudes are very low, and as a consequence, generalizations can be plausibly error-free. However, disciplines such as the humanistic ones are characterized by very high error magnitudes. In addition, the inherent features of these fields sometimes require the use of alternative designs (i.e. quasi-experimental ones), rather than highly controlled experimental ones, which is why inferential developments are so important in our field of work:

the classroom. Below we introduce a synthesis of the most important inferential applications.

One of the core inferential applications is the comparison of the mean execution (DV) depending on the levels of the independent variable. For example, to return to our study of the teaching method, this application would entail a contrast between pairs of means of the type t-test, as we have two levels: control group and experimental group. In general, the ANOVA technique allows us to compare two or more means, whereas the t-test limits the comparison to two means in the levels of the dependent variable. That is why the basic variant t-test is said to be compiled within the ANOVA. That is to say, the t-test is appropriate when we compare two methods, but ANOVA should be used when there are three or more, although we can also use it for two.

Then, depending on the type of sample, we will choose between an independent or dependent type of analysis of variance (ANOVA). The first case, independent, is appropriate for a **between-subject** design, in which the levels of the independent variable are applied to different subgroups, whereas the dependent one is applicable to a **repeated measures –**also called **within-subjects–** design. In this second case all the levels of the independent variable are applied to the same participants. The teaching methods example is a case of between-subject design, because one of the classes receives only one of the levels (IV) of the manipulation –the experimental group: innovative method–, and the other class receives the other level of the variable –control group or traditional method–. That is why we would have a case of independent samples, as comparing the two different groups the measurements in both do not have to be connected. On the other hand, if both methods were implemented in both classrooms, then we would be talking about an within-subject design, because all the participants receive all the levels (values) of the *mani-pulation* (IV). Consequently, observed measurements of the methods compared would correspond to a dependent samples case, as we are comparing the participants' performance as a result of one teaching method (a level of IV) with the same participants' performance in connection with another method (a second level of IV).

Besides finding out, through the framework of the ANOVA, how the levels of the independent variables affect the mean execution in the different 'groups', we may want to *discover* other relationships. Beyond the option of a one-way manipulation, we may have carried out more than one treatment or manipulation, and would like to see the interaction between them, rather than their separate effect. In this case we should use a **Factorial ANOVA**. Imagine a classroom study in which two independent variables appear: the students' knowledge of the L2 (IV A) and the input supplied by the teacher (IV B); the dependent variable would be the degree of communicative competence that changes as an effect of the two manipulated independent variables. The interaction of these two variables means that the effect of the supplied input would depend on the students' knowledge of the L2. Thus, the input would have little effect on the students' communicative competence if their knowledge of the L2 were low; however, input manipulations would affect the degree of competence if the students' knowledge of the L2 were higher. This is an example in which the analysis of the differences in communicative competence should be guided by Factorial ANOVA, given the relevance of the interaction between the two independent variables when they affect the dependent variable, rather than the effects of each isolated individual variable –main effects– have on the resulting change in communicative competence. In other words, it is more interesting to see the interaction of the different factors rather than their individual separate effect.

When experimental designs are more complex than in the examples illustrated above, then data analysis should be carried out through alternative ANOVAs such as **ANCOVA,** which enables us to neutralise the effect of extraneous variables through statistical computations. For example, when researching the differential success of two teaching methods, it may happen that once we have finished our study we discover that the participants' age in the two groups to which we applied the two methods is very different. We then think that age could also be an alternative explanation to the differential success we obtained by the two methods. As we have already carried out our study and we cannot change the design to include the new variable (age) but we can easily obtain its values, we can discard its influence. The way to do this is to incorporate this variable within the statistical technique of ANOVA so that statistical computations neutralise it. This *unexpected* variable we are talking about is called covariate, and the technique we are recommending to use is known as ANCOVA.

A second alternative is **MANOVA,** which enables us to analyse the simultaneous effect of independent variables on more than one dependent variable. On many occasions we make simultaneous measurements of several dependent variables which share many features. For example, two aspects of L2 performance, such as morphology and syntax. Then, we would have to analyse the possible effect of the independent variable (teaching method) on that part common –or shared by– to the two dependent variables. That is why the type of ANOVA to be used here is called MANOVA –Multivariate Analysis of Variance–, which focuses on more than one dependent variable.

A third option is the **Non-parametric ANOVA**, used when mathematical assumptions (for example, errors found in data have to be independent and with equivalent magnitude) in the classic ANOVA have not been fulfilled, as in the case of designs with very few participants (less than ten in each experimental condition). We are not going to describe this type of option because they are too specific and go beyond the aim of this paper. For further details, see Ramos, Catena and Trujillo (2003).

The fourth and last alternative we will consider here is Categorical Data Analysis, with log-linear or log-log statistical computations (where *log* refers to logarithmic compu-tations). This type of analysis should be used when the dependent variable is categorical, or, in other words, non-quantitative. For example, this is the case of variables such as correct answers *vs*. errors –binary– or the notion of communication strategies, which include several categories.

To conclude the section we will focus on describing the statistical alternatives which can be used for those designs which are neither genuine experimental studies nor descriptive ones, but halfway between. First, correlational studies in which a relationship between different sets of variables needs to be established. In this case multiple regression techniques for continuous variables and logistic regression techniques for categorical variables should be used, also based on logarithmic principles. For more specific designs, there exists the alternative of discriminant, canonical and Structural Equation Modelling (SEPATH) techniques. The discriminant ones are used to determine which variables distinguish between two or more naturally occurring groups. The canonical option serves when sets of predicting variables are to be associated with sets of criterion variables. Structural Equation Modelling (Causal Modelling, or Path Analysis) technique hypothesizes causal complex relationships among variables (for example, there are twenty or more predictors for academic success) and tests the causal models with a linear equation system (see Pedhazur and Schmelkin, 1991 for a summary of these techniques).

Finally, somewhere between experimental and descriptive studies we must consider those designs in which the time factor is very important. Case studies (n=1) and interrupted temporal series designs share the aim of quasi-experimental designs, but we measure people on many occasions over a given time, sometimes a very extended period. Let us say we take 10 measurements previous to the manipulation of a teaching method and another 10 after it. We could apply a regression or ANOVA analysis as the ones we described before, but data gathered over time tend to be self-correlated, that is to say, they show a serial dependency pattern. That is why we should resort to special techniques to analyse temporal series, such as the ARIMA models (self-regressive integrated moving average). For details, see Ramos, Catena and Trujillo (2003), and Catena, Ramos and Trujillo (2003).

In accordance with the notion of those designs carried out throughout a given time, in longitudinal designs –i.e. developmental ones– the behaviour of the measures during an extended period has to be observed. That is to say, rather than considering differences resulting from manipulation or treatment, we need to study the growth tendency of those data (see Magnusson, Bergman, Rudinger, and Törestad, 1991) through techniques which model, rather than analyse, the shape of the data. **(note 12)** For example, to describe how L2 performance changes as the age increases.

To conclude, after an enumeration of statistical recipe-like variants which we should be able to identify in order to know where to find that analysis which suits our needs, we come back to our main argument, considered at the beginning of the section. Whatever the statistical variant, we must retain the relative quantity of error we are capable of reducing. This should be our main aim, rather than being able to ascertain whether two groups of students show a significant difference. Perhaps it would be better to isolate the components of our method that best explain the differences and try to be more accurate in later studies.

Works Cited

Arnau, J. (1989): “Metodología de la investigación y diseño”. En J. Arnau y H. Carpintero (coord.) *Historia, teoría y método*. En J. Mayor y J.L. Pinillos (Eds.), *Tratado de Psicología General, Vol. 1*. Madrid: Alambra Universidad. 581-616.

Behrens, J.T. (1997): “Principles and procedures of exploratory data analysis”. *Psychological Methods* 2 (2): 131-160.

Biber, D. (1988): “A typology of English texts”. *Linguistics* 27: 3-43.

Catena, A., M.M. Ramos y H. Trujillo (2003): *Análisis Multivariado: un manual para investigadores*. Madrid: Biblioteca Nueva.

Estes, W.K. (1991): *Statistical Models in Behavioral Research*. Hillsdale, NJ: Erlbaum.

Gass, S. M. (2001): “Innovations in second language research methods”. *Annual Review of Applied Linguistics* 21: 221-232.

Harris, R.J. (1994): *ANOVA: Un Analysis of Variance Primer*. Itasca, IL: F.E. Peacock.

Hatch, E. (1978): “Discourse Analysis and Second Language Acquisition”. In E. Hatch (ed.), *Second Language Acquisition: A Book of Readings*. Rowley, Mass: Newbury House. 403-4.

Judd, C.M. & G.H. McClelland (1989): *Data Analysis: A Model Comparison Approach*. San Diego, CA: Harcout, Brace, Jovanovich.

Judd, C.M., G.H. McClelland and S.E. Culhane (1995): “Data analysis: Continuing issues in the everyday analysis of psychological data”. *Annual Review of Psychology* 46: 433-465.

Juffs, A. (2001): “Psycholinguistically oriented second language research”. *Annual Review of Applied Linguistics* 21: 207-220.

Loftus, G.R. & M.E.J. Masson (1994): “Using confidence intervals in within-subject designs”. *Psychonomic Bulletin & Review* 1(4): 476-490.

Luque, G. (1998): *El discurso oral narrativo en el aprendizaje del idioma inglés en la enseñanza primaria. Implicaciones didácticas*. Doctoral Dissertation. Jaén: Servicio de Publicaciones.

Magnusson, D., L.R. Bergman, G. Rudinger & B. Törestad (eds.) (1991): *Problems and Methods in Longitudinal Research*. New York: Cambridge University Press.

Manly, B. (1992): *The Design and Analysis of Research Studies*. Cambridge: Cambridge University Press.

Maxwell, S.E. & H.D. Delaney (1990): *Designing experiments and analyzing data: a model comparison perspective*. Wadsworth Belmont, California. International student ed.

Nunan, D. (1991): “Methods in second language classroom-oriented research”. *Studies in Second Language Acquisition,* 13: 249-274.

— (1992): *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

Pedhazur, E.J. & L.P. Schmelkin (1991): *Measurement, Design and Analysis: An integrated approach*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Ramos, M.M., A. Catena y H. Trujillo (2003): *Manual de métodos y técnicas de investigación en Ciencias del Comportamiento*. Madrid: Biblioteca Nueva.

Sadish, W.R., T.D. Cook & D.T. Campbell (2001): *Experimental and Quasi-experimental Designs for Generalized Causal Inferences.* Berkeley: Houghton Mifflin.

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

Van Lier, L. (1988): The Classroom and the Language Learner. London: Longman.

1. As when you find out you should have included a control group, you haven't and while it took you three years to carry out the research task at hand, your doctoral dissertation deadline is approaching and how can you solve this problem?

2. If a teacher/researcher wanted to study the effect of age on second language acquisition, he/she would be at pains to gather all the information and data coming from this field, as a lot of research has been done on the topic and there is an enormous amount of accepted results. On the other hand, this teacher would also be at pains to find out information about –for example– the deaf language of two students, but this time because of the scarceness of the data. Either situation determines later steps and research strategies.

3. Reproduced again on different occasions, with different participants and situations to obtain similar results (Sadish, Cook & Campbell, 2001).

4. I.e. The existence of kernel sentences and transformations by way of general and universal transformational rules.

5. It has in fact been contested by later models, which do not attach such an important role to LAD and UG transformation, believing they are not as universal as they seem. The theory has answered by creating an alternative model with alternative principles (i.e. Minimal Distance Principle).

6. In Luque, (1998). *El discurso oral narrativo en el aprendizaje del idioma inglés en la enseñanza primaria. Implicaciones didácticas*. Jaén: Servicio de Publicaciones.

7. Note that the lack of a theory would mean that the research is not correct.

8. In fact, controlled but artificial situations tend to be associated to internal validity, whereas external validity can be related to situations which are not-controlled but rather less artificial (see Sadish, Cook & Campbell, 2001 for a revision of this topic).

9. What evidence (data) will be collected and what procedure will be chosen to gather those data are aspects that depend on the type of research strategy, problem and theory the teacher/researcher has favoured. They also depend on the operative hypotheses and variables to be verified, although the teacher/researcher has several options here: he/she may use or adapt a ready-made collection procedure, or create one starting from scratch. For a recent revision of innovations in L2 research methods, see Gass, (2001).

10. There is always a degree of interpretation of evidence, regardless of the type of research carried out.

11. There are different options included in this model, which range from a single very concrete prediction to a very sophisticated one within an elaborate mathematical model.

12. Among others, we can mention Linear and Polynomial regression (i.e. quadratic), the logarithmic function or the exponential function (linearizables through simple mathematical transformations of the raw data) and Intrinsically Nonlinear Regression Models, such as the General Growth Model and the Probit-Logit-General Logistic Regression Model (for Categorical variables) and ARIMA.