

# Data Collection in SLA Research

Thomas H. Goetz<sup>1</sup>

INTRODUCTION

THE QUALITATIVELY HUED METHODS

THE QUANTITATIVELY HUED METHODS

APPLICATIONS

CONCLUSION

## INTRODUCTION

This research note is done out of a motivation to bring into focus the various methods for data collection in SLA research. They fall into two broad paradigms: qualitative and quantitative. These represent more than two relative extremes but terminals on the Qualitative-Quantitative Continuum of Research Methodologies. A qualitative paradigm concerns understanding the acquisition of a second language from the learner's frame of reference. Data are collected from naturalistic settings and there is no control group. Therefore it is highly subjective in nature. The researcher can be a participant right along with the subjects and there can be considerable exchange between the two. This paradigm is by design discovery oriented and exploratory. In SLA the longitudinal approach is one example where the data of language acquisition are described holistically as process oriented that is real, rich, deep and dynamic.

---

<sup>1</sup> I would like to thank Drs. Gladys Valcourt of Temple University's Osaka Branch and Steven Ross of Kwansai Gakuin University, Sanda Campus for their help in proof reading earlier versions of this report.

Adding to its subjectivity, these are often single case studies and results are ungeneralizable.

On the other side is the Quantitative Paradigm; it seeks to quantify the facts or causes behind facts of human or social behavior objectively with minimal emphasis on the role of individuals or conditions that are beyond the ken or the researcher's control. In this way the researcher is removed from the data and willingly maintains an outsiders' perspective. In SLA, the cross-sectional approach is typical of this paradigm in that they are verification-oriented. Researchers seek to confirm particular hypothesis which can be inferred for further research: generalization. Cross-sectional studies are outcome or product oriented. Quantitative research requires reliable or hard data that can be replicated in other stable situations.

The above mentioned represents two major paradigms which by themselves are not mutually exclusive, rather one blends into the other over the qualitative-quantitative continuum of SLA research methodologies. Larsen-Freeman and Long list seven research methodologies in a Likert-type scale manner ranging from purely qualitative to purely quantitative. Those methodologies are: Introspection, Participant Observation, Non-participant Observation, Focused Description, Pre-experimental, Quasi-experimental, and Experimental.

## THE QUALITATIVELY HUED METHODS

Introspection can be described as purely qualitative in that learners tell researchers what is going on in their minds and in their own behavior for insights into their acquisition (or in-acquisition) of a second language. Its advantages are that it is highly intuitive and it has been used widely in the field of psychology. Its disadvantages range from the nature of the data itself (is self-reported data really

data?) to wondering if what the learners report is really what is happening. Introspection appears often in SLA studies that seek to describe affective factors associated with a learner's acquisition or lack thereof: interlanguage studies, attitudes and motivation and their contribution to fossilization are a few examples.

Participant Observation, as its name implies, involves the researcher in the activities under study. Usually there is no pre-set hypothesis; its very exploratory. The ethnographic study, where the life and culture of a society or ethnic group gets studied by personal or participant observation, would be an example of this research method. A typical L1 and or bilingual example is that of recording everything a child says when she was acquiring her language or languages. Its advantages are that it is often longitudinal, the participant and the observer often talk and interact. The disadvantages of participant observation limit its use in that the data collected may not be as natural and dynamic as assumed. The mere presence of an observer can bring about what's known as the Hawthorne Effect:<sup>2</sup> that the subject will out-perform herself knowing that she is being monitored by someone "important." Moreover, this approach works well with subjects whose personalities predispose them as being articulate in the first place; with these limitations in mind, participant observation provides a good framework for long-term language acquisition studies. After the data have been collected over a span, for example, the acquisition of grammatical morphemes, they can be categorized and processed according to an implicational scale, thus showing the order in which such morphemes are acquired. This is a good starting point for more carefully controlled studies especially in interlanguage and the acquisition of grammatical structures in either developmental or variational sequences.

---

<sup>2</sup> Please refer to the Appendix for a list of salient terms that apply to SLA research.

Non-participant Observation is a method in which the researcher is not noticeably around. One-way mirrors, hidden microphones and or video cameras typify this approach. Data are processed after collection which is different from participant observation where processing often happens during collection periods through researcher—subject interaction. It is longitudinal but usually involves more than one subject. Its advantages are found in the relative “strength” of the data in that participants may not know they are being monitored; Hawthorne Effect may not apply. The data are rich in that the researcher can comprehensively detail the utterances of the subject(s) and while no hypothesis is being tested, the participant and non-participant studies can generate hypotheses for further research. Its disadvantages center around the validity of the data collected and the absence of any independent variables; there is a risk factor of “reading into” the data by the researcher that which may not be there at all. Its usefulness is limited to descriptive studies and no one should generalize anything from such a subjective research method such as this.

Focused Description can be described as favorably different from the first three methods mentioned in that the researcher isolates variables for observation. As the name implies, it focuses on something particular descriptive purposes often through interaction. The researcher wants to see how, in an L2 setting if certain behaviors are present or at what stage learners are, morphologically, on the inter-language continuum. The measures have pre-set categories which the learners may or may not meet and the researcher keeps a record of the number of correct uses which can later be converted into a percentage. The Bilingual Syntax Measure by Dulay and Burt is one example of a focused descriptive study. When participants reach a certain percentage of proficiency with a particular morpheme, then that morpheme can be described as having been acquired. This methodology gains much strength in that it can

demonstrate the extent to which certain morphemes are acquired and it can also be applied to correlational studies.

Correlational studies seek to determine if there exists a relationship between two variables such as learner's motivation and second language proficiency (see Gardner and Lambert: 1972). Its advantages are found in its ability to narrow down that which researchers wish to measure so the data can speak more for itself. Focus descriptive studies do not require a lot of time nor resources when compared against the three above. With more subjects being evaluated, the findings are relatively more concrete than the open-ended studies. Researchers can, albeit cautiously, generalize somewhat about their findings. The disadvantages of the focused descriptive studies center around the problem that with its limited focus, say on the acquisition of morphemes, these kinds of studies may overlook the multi-dimensionality of second language acquisition. Morpheme acquisition is not the only way to assess language proficiency. Good sociolinguistic competence and discourse ability can often provide that needed edge of compensation for a learner who may not have all the morphemes yet; that learner will be a better communicator. Another disadvantage which can limit the range of use for the focus descriptive study is the persistence of effect. At a particular time and place, a particular learner may test well, but that does not always assure that the ability shown is the learner's base level ability.

## THE QUANTITATIVELY HUED METHODS

Pre-experimental can be described as something less than a true experiment in that it lacks a control group and participants are not randomly selected; groups are intact, or pre-selected. In a true experiment researchers aim to make statements about causality between two or more variables; the pre-experiment by design cannot provide the evidence to show causality but it does allow for a trying

out a new method. If there was some effect, then they can try again with a more rigorous experimental design. Often it looks like this  $X_1 \text{ T } X_2$  where  $X_1$  is a test or battery of tests whereupon subjects get some kind of treatment T, intensive language study perhaps, and then there is a final assessment,  $X_2$ . Its advantages are in its sheer simplicity of design but that is, however, its greatest disadvantage. Its range of use is best limited to the collection of "soft" or data concerning constructs that by their nature are hard to define and assess such as attitudes and motivation in second language acquisition.

Quasi-experimental is design in which causation is sought out; control groups are constructed, but the groups are not randomly assigned. The independent variable is the one that is manipulated and the dependent variable is the outcome; results are then compared against the performance of the control group. Its schematic design looks like this: for the experimental group —  $X_1 \text{ T}_a \text{ } X_2$  but the control group looks like this  $X_1 \text{ } \phi \text{ } X_2$  or  $\text{T}_b \text{ } X_2$ . There are many variations to this basic design. One of the most common is the design for persistence of effect in which a third (or more) assessment(s) is (are) administered at a preset interval(s) to see how subjects remember the treatment. Its advantages: it is very close to a true experiment and SLA teachers can carry out experiments with their classes because each class is an intact group; very practical. Its disadvantages are found when researchers try to generalize. Since groups are not randomly assigned, there are severe restraints on findings. This design is also subject to the histories of its participants. A researcher may not know that a third of the students in the experimental group graduated from English speaking high schools; this history factor would undermine the experiment's internal validity. This limits its range of use due to lack of randomization.

Lastly Experimental can be described as an ideal situation in which all factors are held constant. This is much easier to do in

bio-medical research than in the social sciences, SLA is no exception. There are at least two groups in a study, a control group and a treatment group, and the subjects are randomly assigned. Having randomly assigned groups adds internal validity to the experimental design in that it assures that the results were derived from at least two groups that were similar if not almost equal to begin with. Its basic experimental design is that of the quasi-experimental design. Its advantages are powerful in that if a difference can be measured between two or more groups, then there is a causal relationship between treatment and outcome; there is no shortage of statistical procedures by which raw data may be processed by to determine their significance. If research is done in this way, SLA researchers can be bold in generalizing their findings to the ESL/EFL world. Its disadvantages involve removing from a naturally occurring context the variable which needs to be assessed. In contrast to the real world, the true experimental design has to maintain a relatively sterile research environment. This is often seen as simplification and just plain "unnatural." Sometimes experimental conditions cannot be met. When that happens, the quasi-experimental design is most appropriate; it is able to account for idiosyncrasies of human behavior.

When doing research on the interlanguage of requests, I would set up my action research with a discourse completion task and proceed with the quasi-experimental method. Where I work I cannot easily randomly select students. The request by itself is a construct that, by its nature, lends itself well to discourse completion and comparison against a control group (native speakers of English and the subjects' L1(s)). I would also include an opportunity for introspection by the subjects to tell me why they performed the way they did; after all each subjects' interlanguage is on the one hand particular (hence the introspection), but on the other hand, it is a shared phenomenon (which can be ordered and sequenced). Beebe,

Takahashi, and Uliss-Weltz used the quasi-experimental design in their study on pragmatic transfer in ESL refusals; they did not say that the subjects were randomly selected. Given my situation, it appears to be a useful method.

## APPLICATIONS

When doing research on, for example, the emergence of English articles, the focused descriptive method of data collection would be most useful. Subjects need only to be given a measure, something like Dulay and Burt's Bilingual Syntax Measure. It looks like a one shot design in which we assume that  $T_a$  stands for measure and  $X_1$  stands for the measure's results. Its main strength is found in its simplicity — the teaching of article system followed up by a test. Its greatest weakness is that it has no way of controlling or moderating intervening variables. Therefore one cannot reliably conclude much due to the possibility of threats to validity and reliability of the research, but it does provide useful descriptive data which can provide impetus for further research. It also lends itself well to implicational scaling and scalograms which are highly visual and reliable.

When doing research on a comparison of Ns verses NNs teachers' evaluations of error, the qualitative data collection methods non-participant observation and introspection would be most useful, as this is more ethnographic. Observations as they naturally occur, discussions between native and non-native teachers would be the data. Such data would then be used to form a hypothesis or hypotheses. This requires a microanalysis of the teachers' interactions by: (1) recording their interactions on video followed by (2) consultation sessions where the teachers would be shown their segments of error evaluation for analysis frame by frame in an ideas exchange meeting with me. There they could tell me their rational behind their evalua-



tions. The findings would not be not generalizable, but they would serve to provide a local dynamic and impetus for further work.

## CONCLUSION

The purpose of this report has been to discuss the major areas in which data are collected in the field of Second Language Research. In fact, in light of the borrowing from other areas within Applied Linguistics and psychology, a heightened awareness of data limitations within the categories discussed in and of itself can give strength to the overall work of a researcher and ultimately the field. Simply having collected a data set and running “some kind of stats” on it does not a contribution make. Setting the data within its appropriate strata does.

## Appendix A

Salient terms used in discussing research.

**CONCURRENT VALIDITY:** relates to the instrumentation.

**CONSTRUCT VALIDITY:** refers to whether or not data are being interpreted validly.

**CONTENT VALIDITY:** refers to the extent to which an inventory's questions are related to the construct.

**CONVERGENT VALIDITY:** This refers to a situation in which a battery of tests measures its method instead of an isolated trait.

**CORRELATIONAL EVIDENCE:** This refers to construct validity. When creating instrument measures for listening skills, we need to isolate the trait we are assessing. We need a convergence of two or more different measures of the same trait.

**COVARIANTS:** as students' extracurricular activities vary, so do their TOFEL scores. This is a way of moderating control variables.

**COVARIATE:** describes a state of correlation between variables.

DEPENDENT VARIABLE: This is the result of the experiment.

DESIGN: either internal or external. Internal concerns instruments and participants. Instruments include internal consistency.

ERROR: This refers to individual differences that is observable which does not reflect any true difference between them. The sources of individual differences that effect the dependent or the independent or the relationship between the dependent and the independent variable. It also amounts to unwanted influences. Error can be random or systematic.

EX POST FACTO DESIGN: This refers to a design that shows a relationship or a correlation between two or more variables typically over time. The independent variable is the predictor and the dependent variable is the outcome.

EXPERIMENTAL DESIGN: This refers to a design in which causation is sought out. The independent variable is the one that is manipulated. The text refers to this as "T" for treatment. The dependent variable is the outcome.

FIXED VARIABLES: are those which are limited to set variables within the test design. Among East Asian languages, Chinese, Korean, and Japanese there are three fixed nominal variables.

HAWTHORNE EFFECT: This refers a phenomena when a subject becomes aware that they are in an experiment, eg. an observer is present and so the subject tries extra hard.

HISTORY: This applies to individual students in particular classes and the causal relationships in the methodological gain in proficiency. The socioeconomic status of participants in an experiment can be a factor.

HISTORY FACTOR: participants come from different backgrounds and they are either enriched or lacking.

INDEPENDENT VARIABLE: this is manipulated and the outcome is the dependent variable. An extra-variable may be correlated as well and in so doing possibly reduce the margin of error.

**INDIVIDUAL DIFFERENCES:** This can be the normal features that any two populations may have.

**INSTRUMENTATION:** This refers to the ways in which we form our measures or forms of assessment. All of our tests contain some degree of error.

**INTERNAL CONSISTENCY:** This refers to asking one question more than once. Positively worded and negatively worded. If there's no consistency, perhaps the participants really don't understand the propositional content. It pertains to all tests.

**JOHN HENRY EFFECT:** This is when the control group becomes aware that they are the control and they chose to beat the experimental group especially when the experiential group is unaware. It happens with children most often.

**MATURATION:** This is the process of people getting better with the process of time. Anyway, through other sources of viable input regardless of treatment, learners learn; the control group does not work.

**METHOD ARTIFACTS:** refer to a style of test and the way it is made and the way it is administered. In all these errors can offset the data.

**MORTALITY:** people drop out and quit from an experiment. This can severely effect reserach outcomes.

**OPERATIONAL VARIABLES:** keep us interested in groups, conditions of learning, methods of teaching, variables of interest such as age, native language, literacy in the L1 and its script. There are many more, cognitive dependence and independence, aptitude, motivation.

**PARTICIPANTS:** can be effected by several ways: self selection or volunteers they can offset the internal validity of the project.

**PRE-TEST DIFFERENCES:** When groups are compared and they are different before the independent variable of inference is applied, this would flaw the internal validity of the experiment. To over-

come pretest differences is to give a pretest and this will overcome the differences.

**PREDICTIVE VALIDITY:** refers to what extent the researcher can guess what become of the data.

**REALISM:** is some unconstructed reality, as in physics. There is a real phenomena.

**REGRESSION ANALYSIS:** removes error that is in the design.

**RELATIVISM:** This presupposes that humanity has no ultimate truth, theory, construct, or abstraction. Anxiety is a construct. Everything we do is constructed in theories. Relativism promotes argumentation, we seek relative interpretations.

**REPLICATION:** This refers to an experiment's ability to confirm itself in other laboratories.

**RESEARCH DESIGN:** This requires an internal validity of the operational variables themselves.

**RESENTFUL DEMORALIZATION:** This happens when the experimental method of interest is one which involves high technology but only a limited number of students can participate. The control group learns that for experimental purposes they can't access the new lab. They give up.

**SELECTION OF PARTICIPANTS:** Volunteers in the experiment and control are not biased from the outset.

**SELF-FLATTERY:** People often see themselves in an ideal light in a large institution. The instrument itself creates some error in that it is not stable from day to day.

**TASK COMPLEXITY:** refers to a situation say instructions that are more difficult than the inventory itself.

**TEST EFFECT:** This refers to a student who responds to the fact that he is aware of his participation and alters his frame of mind for the benefit of simply taking a test.

**TRIANGULATION:** Refers to the interaction in describing the same phenomenon among an observer, students, and teacher. One can

arrive at some convergent evidence that would go with internally consistent survey.

UNRELIABILITY: This is the biggest threat to experimentation in that the scores obtain today would be the same the next day. Attitude and motivation are very latent and therefore subjective constructs.

VARIABLES: These are the phenomena that we construct and name because there is something in the experience of the researcher of interest: motivation-continuum, age-continuum but is it often an ordinal scale (15-20, 21-25), sex is a non-arithmetic dichotomy.

[Works cited]

- Burt, M., Dulay, H., & Hernandez, E. (1973). *Bilingual Syntax Measure*. New York, NY: Harcourt Brace Jovanovich.
- Gardner, R. C. and W. E. Lambert (1972) *Attitudes and motivation in second language learning*, Rowley: Nwbury House
- Janet T. Spence, J. W. Cotton, Benton J. Underwood, Carl P. Duncan (1976). *Elementary Statistics* (third ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Lazaraton, A., Hatch, E., (1991). *The Research Manual*. Boston: Heinle & Heinle Publishers.