Annual Review of Applied Linguistics (2001) **21**, 221 232. Printed in the USA.
Copyright © 2001 Cambridge University Press 0267-1905/01 $9.50

13. INNOVATIONS IN SECOND LANGUAGE RESEARCH METHODS

Susan M. Gass

Acceptance of the claims made by researchers in any field depends in large part on the appropriateness of the methods used to gather data. In this chapter I focus on two approaches to research in second language acquisition: (a) various types of acceptability judgments or probes aimed at assessing acquisition of syntactic structure; and (b) various types of stimulated recall designed to gather learners accounts of their own thought processes. Both methods attempt to overcome a principal problem in psycholinguistics: the desire to describe a learner s knowledge about a language based on the incomplete evidence stemming from learner production. Refinements in acceptability judgments have come from some newer multiple-choice or truth-value story tasks that allow researchers to determine the level of learner knowledge about particular syntactic structures (in the examples here, reflexives). Stimulated recall offers some additional perspectives, but its usefulness can be greatly affected by the temporal proximity of the recall to the original task; the amount of support provided to prompt the recall; and the nature and amount of training given to both interviewer and interviewee. While these newer research methods can improve the accuracy and variety of data available to SLA investigators, research methods drawn from L1 acquisition or L1 research cannot necessarily be assumed to be equally valid when used to examine L2 acquisition.

The topic of research methods is a vexed one. Gathering data is often limited only by one s imagination, yet, at the same time, the method itself must be valid and reliable. Thus, one is faced with a constant dilemma of how to elicit data using methods that are sufficiently understood, and hence, have face validity, while avoiding data that, because of the methodology used, are ambiguous in their interpretation. In fact, consensus about the validity of any field of research is dependent on an understanding of the methods used to gather data. As a reflection

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

of the uncertainty of what methods in second language acquisition (SLA) and applied linguistics research stand for, one needs only think about the number of books and articles reflecting this topic (cf., Davis & Lazaraton, 1995; Gass & Mackey, 2000; Han, 2000; Markee, 2000; Schachter & Gass, 1996; Tarone, Gass, & Cohen, 1994; Yule, 1997).

As a way of illustrating the complexity of issues that come with newer methodologies, in this chapter I focus on two distinct areas of second language research, one dealing with the acquisition of a particular syntactic structure (reflexives) and the other dealing with a methodology that is being used of late to illustrate thought processes of second language learners. The field of SLA is multidisciplinary. In its history as well as in its present form it is influenced by numerous source fields. This has led to many discussions of methodology inasmuch as each source discipline brings with it its own favored research traditions and prejudices (both positive and negative) about the validity and insufficiency of certain methods.

As an example, consider linguistically-based SLA research. Most research within linguistics has traditionally been conducted using a form of acceptability judgment.[1] This was clearly the case when SLA research had its beginnings in the 1960s and is also the case, although clearly to a lesser extent, in today's research climate (see Juffs, this volume). From the early years of SLA research, it became apparent that without an examination of what information acceptability judgments were tapping, one could not accept their use unquestionably. This was clearly recognized when, even while using acceptability judgments as a means of gathering data, researchers often felt the need to justify their use. In addition, over the years, their validity and reliability have been called into question using empirical data as opposed to theoretical argumentation alone as evidence (Cowan & Hatasa, 1994; Ellis, 1990; 1991; Goss, Ying-Hua, & Lantolf, 1994). Even those who believe that acceptability judgments are a valid and reliable means of collecting data (e.g., Gass, 1994), recognize the difficulties involved.[2]

Some of the early debates about methods in second language research centered precisely on this form of data elicitation. Selinker (1972, p. 213) ignited the debate when he stated that researchers should "focus ... analytical attention upon the only observable data to which we can relate theoretical predictions: the utterances which are produced when the learner attempts to say sentences of a TL." This view, which clearly denounces acceptability judgments as well as any sort of introspection as a viable data source, given that judgments are not "produced utterances," had supporters and detractors. Corder (1973), for example, pointed out that spontaneously produced utterances provide only a part of the picture. If one wants to obtain information about the knowledge that learners have, then one must also have a means to determine which sentences they think are possible in a second language (i.e., grammatical) and which they believe are not possible in a second language (i.e., ungrammatical). To accomplish this, some

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

means other than what learners produce is necessary (see additional discussion in Gass, 1997).

Once acceptability judgments were no longer used as the quintessential data-elicitation method for obtaining information about second language learners knowledge of the L2, other sources of data became more prominent, each with differential claims of validity and reliability (e.g., elicited imitation, Bley-Vroman & Chaudron, 1994; Munnich, Flynn, & Martohardjono, 1994; magnitude estimation, Bard, Robertson, & Sorace, 1996; sentence matching, Gass, in press; Plough & Gass, 1999; Gass, in press). These and other newer methods have gained credence over the years although none is without problems.

The Evolution of Methods in Investigating Reflexives

The preceding discussion has focused primarily on one method of second language data elicitation that has engendered controversy since the 1960s. We turn now to a discussion of one particular grammatical structure and focus on the methodological issues surrounding an understanding of how that structure is acquired by second language learners. This discussion is intended to demonstrate the evolution of methodology, as one data-elicitation method superceded another in order to hone in on a way of gathering data that appropriately reflected learners knowledge of a second language. The controversy elucidates the validity problems discussed earlier.

The issue at hand is the acquisition of reflexives. The particular issue concerns sentences such as the following:

1. Sally told Jane to wash herself.

The question relates to the referent of *herself*. In English the interpretation of *herself* is limited to Jane. In other languages, however, this sentence may be ambiguous. The acquisition question is: How do learners learn to restrict or expand the interpretation of reflexives when their native language differs from the target language? Complexities involve the number of clauses in a sentence (monoclausal vs. biclausal sentences) and, in biclausal sentences, whether the embedded clause is finite or nonfinite. Sentence (1) illustrates a monoclausal sentence and sentences (2) and (3) illustrate biclausal sentences (finite and nonfinite).

2. Biclausal, finite embedded clause (from Glew, 1998)
   Tom noticed that Bill was looking at himself in the mirror.

3. Biclausal, nonfinite embedded clause (from Glew, 1998)
   Tom told Bill to trust himself more.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

The answer to the acquisition question is highly dependent on methodology. For example, one early study (Finer & Broselow, 1986) used a picture-identification task whereby learners were given a sentence containing a reflexive and were shown two pictures. The task (see also Eckman, 1994) involved selecting a picture that accurately reflected the meaning of the sentence. Lakshmanan and Teranishi (1994) pointed out that the picture-identification task itself may be questionable because a learner may be able to select the correct picture based on nongrammatical knowledge, and hence, the results may not be valid at all. Others (e.g., Hirakawa, 1990; Thomas, 1989) used a multiple-choice test as a means of elicitation. For example, learners were given a sentence such as the one given in (1) and were asked whom *herself* could refer to. Choices, such as those in (4) were offered.

4. (a) Sally
   (b) Jane
   (c) Either Sally or Jane
   (d) Someone else
   (e) Don t know

Lakshmanan and Teranishi (1994) noted that this method for gathering information about learners knowledge of reflexives is flawed because it provides information only about what *can* be a possible antecedent of *herself*, but not about what *cannot* be a possible antecedent. Glew (1998) similarly noted that in doing either a picture-identification task or a multiple-choice task, a learner may have a strong preference for a sentence. In such cases, that individual may select only one of the choices and not consider all the possibilities of interpretation. Thus, one is left not knowing what a nonresponse means. Does it mean that the learner did not consider all possibilities or that she or he did consider all possibilities and that the sentences that were not selected are ungrammatical for that learner? White, Bruhn-Garavito, Kawasaki, Pater, and Prévost (1997, p. 148) pointed out that the fact that they [learners] choose only one interpretation does not necessarily mean that the other is excluded from their grammar. It is essential to recognize that both grammatical and ungrammatical information are necessary to completely understand what an individual knows about a second language. As Glew (1998) pointed out, traditional methods tap preferences rather than complete intuitions.

Another difficulty in dealing with reflexives has to do with interpretation based on contextual or pragmatic information. A sentence such as (5) generally elicits a response such that the typical interpretation from studies of native and nonnative speakers of English is that *actor* is the antecedent of *himself* (White et al., 1997).

5. The actor gave the man a picture of himself.

This is so for grammatical and pragmatic reasons (actors are more likely to distribute pictures of themselves rather than distribute pictures of others).

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

However, a reading whereby the second NP is the antecedent is also possible. For example, consider the following situation: A man had had a lot of pictures taken of himself. An actor found one of these pictures and The actor gave the man a picture of himself. As noted above, there is a difference between preferences and what is allowed and disallowed. Sentence (5) in its most typical preferred reading yields *the actor* as the antecedent of *himself*. However, that does not tap the full range of possibilities because, given a different context, as mentioned above, *the man* is also possible. In other words, a learner may select *actor* because that is the obvious and most usual reading, while not considering other possibilities.

White et al. (1997) used a truth-value storytelling task to elicit data. The context consisted of a short description of a context, followed by a sentence. Participants were asked to state whether the sentence accurately described the context. An example from White et al. (p. 153) is given in (6).

6. Susan wanted a job in a hospital. A nurse interviewed Susan for the job. The nurse asked Susan about her experience, her education and whether she got on well with people.
   The nurse asked Susan about herself. True/False

With a variety of contextual clues, one can distinguish between preferences and intuitions. White et al. also gave a picture-identification task to their same group of learners. In this task, the participants read a sentence and then had to state whether that sentence did or did not match a picture. In comparing the results from the two tasks, they found that the story task was better at showing the allowability of objects as antecedents as in (6) above, where *Susan* is the antecedent of *herself*. However, to further show the complexity of adopting any method, they noted that when participants did a picture-identification task, they apparently first read the sentence and then looked at the picture. Their first impression may have blocked future interpretations, much as the first viewing of an optical illusion may preclude us from seeing other interpretations of a figure.

Truth-value story tasks provide a way of manipulating context to demonstrate the effect of pragmatics on judgments of acceptability. For example, the sentence in (7) below (Glew, 1998), would lead one to an ungrammatical conclusion (in English) that *himself* referred to *doctor*. The story (and common sense) tells us otherwise. If a learner selects True (assuming an understanding of the story), one can assume that a learner allows reflexives to be coreferential with the subject of the matrix sentence; if a learner, on the other hand, selects False, the learner understands that *himself* refers to doctor and this is false, given the context. Thus, through this method, one can determine the limits of ungrammaticality for learners.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

7. John was sick in the hospital. He was nervous because he needed an operation. A doctor came into John s hospital room and said, Hello John, I am the doctor. I will operate on you tomorrow.
John heard that the doctor will operate on himself. True/False

In yet another variation of data elicitation, Lakshmanan and Teranishi (1994) presented the target sentence with all of the nonpossibilities for the reflexive antecedent. An example appears in (8).

8. John said that Bill saw himself in the mirror.
   a. Himself cannot be John. agree disagree
   b. Himself cannot be Bill. agree disagree

In this way, all of the possible responses can be ascertained.

The preceding discussion has illustrated the evolving sophistication of and concern with research methodology in order to establish the validity of the methodology itself and ensure that the data elicited reflect what researchers believe they reflect. In the next section, I deal with stimulated recall, a form of introspection.

Stimulated Recall

The goal of SLA research is to determine what second language learners know about a second language (i.e., what sorts of grammars are formed and are not formed), when they come to know it, and how they come to know it. Because traditional methodologies have been in debate due particularly to their face validity, new methods have come into greater use and greater acceptance. One that has received attention of late is recall methodology, particularly stimulated recall methodology.

SLA research, like all psycholinguistic research, is faced with the inevitable problem of needing to determine the processes involved in learning, yet not being able to observe those processes. All that is observable is what a learner produces, in writing or in speech or in response to specific researcher probing. Researchers in second language acquisition have over the years developed greater sophistication in the techniques used for probing, with the goal of determining underlying linguistic knowledge and/or linguistic processing. Many are borrowed and/or adapted from other fields.

Various methods have been used to determine underlying linguistic knowledge, including having learners introspect about their knowledge. Like all methodological tools, introspection (acceptability judgments being one type) has not been without its detractors, but it is now being used once again with some frequency and with increased confidence. In this section I consider the broader area of verbal reporting with a particular focus on stimulated recall.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

As mentioned above, introspection, of which stimulated recall is one part, has not always been accepted as a valid tool for gathering information about knowledge of language (first or second)[3] (see also Smagorinsky, this volume). In fact, during the behaviorist era, valid data were those that were *produced*; verbal reporting, a form of introspection, was not an allowable source of information (see Gass & Mackey, 2000 for an extensive discussion of the history of introspection).

In general, verbal reporting is a data-gathering method whereby individuals are given a task (e.g., writing a composition or solving a problem) and asked to say what is going through their minds as they are working their way through the task. Clearly, it is not possible for there to be simultaneous verbalization at all times. This is particularly the case when reflection concerns an oral interaction. In those instances, verbal protocols are collected after the event. There are numerous ways in which this can be done and numerous factors that must be considered in determining the validity of the verbalized thoughts as an accurate reflection of the actual thoughts at the time of the event.

When dealing with stimulated recall, the verbalization is done with some support. The support may be in the form of a written piece that the learner has produced or a video or audio of some event (e.g., an oral interview or, in the case of second language interaction research, a task involving an oral interaction). An important aspect of stimulated recall is the temporal relationship of the recall to the original event. Recall can be consecutive (immediately following a language event, see Mackey, Gass, & McDonough, 2000), delayed (perhaps a day or so after the event) or nonrecent (e.g., research in which one reflects over a period of time about learning strategies). Bloom (1954) found decreasing accuracy as a function of the time interval between the recall (whether stimulated or not) and the event being recalled. In other words, with greater time delays, it is not clear what can be claimed with regard to the memories that are being accessed.

There are numerous ways that stimulated recall can fail, some of which have been mentioned above. In general, not only does one have to set up an appropriate structure (e.g., time interval, strong support stimulus), but one also has to adequately train the participants (both interviewer and interviewee) in the art of verbalization; furthermore, one has to avoid the common pitfall of asking the wrong question during a recall procedure. In particular, it is often difficult to separate a reflection relating to the moment of recall from a reflection about a previous event. For example, consider the following exchange between a researcher and a second language learner. The learner had participated in a "Spot the Difference" task in which an interviewer had frequently stopped her to seek clarification. Following the initial episode, which had been videotaped, the researcher replayed the tape, asking the learner what was going through her mind at the moment of the original episode (from Gass & Mackey, 2000, p. 91). The first three columns reflect the original episode and the second two are the comments from the stimulated recall episode.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

NNS Original Comments	NS-1 Feedback	NNS Response	NS-2 Stimulated Recall Prompts	NNS Response
He just look at-look at me				
	He s what			
		Look at me		
	He s looking at you?			
		Yeah		
			Why was she saying that?	
				It s a um, I cannot cannot cannot answer clearly
			What do you mean clearly?	
				I have to answer, I have to answer correct because the word is so weird
			You are saying the word is weird?	

In this instance, the interviewer did not focus on the event, but rather focused on the moment of the interview (which immediately followed the event). She used the present tense when she said What do you mean clearly? and later You are saying the word is weird? It was not made clear by the interviewer that the learner (interviewee) was supposed to be recalling what she was thinking at the time of the event itself. Results such as these are, of course, suspect.[4]

In sum, stimulated recall methodology is a relative newcomer to the data-collection repertoire of second language research. It provides a window onto the thought processes of learning, but must be used with care if the results are to be valid.

Conclusion and a Cautionary Note

Innovation in research methods comes with a price tag attached. Many researchers have become dissatisfied with traditional SLA methods and have turned to methods that have been used in other fields. These tend to come more from psycholinguistics than from linguistics. Among them are reaction time experiments

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

and sentence matching. But, even though the methods are widely accepted in other fields, they must be tested in the SLA arena. For example, sentence matching tasks have been used recently in SLA studies, yet an examination of their validity shows them not to give the same information for nonnative speakers as they do for native speakers (Gass, in press; Plough & Gass, 1999). Thus, innovation must be careful and deliberate. One cannot necessarily come to the same conclusions about nonnative speaker knowledge as one can about native speaker knowledge. Methods do not always transfer.

Notes

1. We differentiate between acceptability judgments and grammaticality judgments even though the latter term is commonly used for both. Technically, when one conducts research, one asks someone if a particular sentence is acceptable. From this we infer whether the given sentence is grammatical (that is, whether it can be generated by the grammar).

2. A difficulty in using acceptability judgments with second language learners as opposed to native speakers of a language has to do with the type of knowledge that second language learners have. When dealing with second language data as opposed to native speaker data, learners do not have total control over the area of grammar that they are being asked about. When a native speaker of English is asked if the sentence *That s the woman I talked about her* is a possible sentence or not, we can safely assume that a negative response means that the sentence is not part of the grammar of English. But, when we ask a second language learner about the same sentence, it is not clear what to make of a negative response. Does the negative response arise because the sentence is truly ungrammatical in this individual s second language grammar or did the learner have no idea and guessed? The latter reflects indeterminate knowledge knowledge over which there is no information on which to base a judgment.

3. Second language is to be interpreted broadly and refers to any language learning after the first (i.e., second, third, fourth, and so forth).

4. These data were originally collected for a study dealing with interactional feedback (Mackey, Gass, & McDonough, 2000). Because of the problems inherent in this interview, these data were eliminated from the final database.

ANNOTATED BIBLIOGRAPHY

Davis, K. & Lazaraton, A. (1995). (Eds.). Qualitative research in ESOL [Special issue]. *TESOL Quarterly, 29*, 3.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

This is an edited special issue of the *TESOL Quarterly*. The articles include treatments of qualitative research in general as well as a number of ethnographic studies of classroom research. There is also a generic discussion of the role of qualitative research (mostly ethnography) in teacher education.

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

This book deals with verbal reporting in general with a specific focus on stimulated recall. It treats the topic from both an historical perspective and a how to perspective, providing the reader with a detailed discussion of ways to appropriately conduct a stimulated recall procedure. The presentation includes a discussion of reliability and validity as well as a treatment of pitfalls to avoid. The book closes with suggestions for studies that would benefit from stimulated recall procedures to address follow-up questions arising from hypothesis-generated research.

Glew, M. (1998). The acquisition of reflexive pronouns among adult learners of English. Unpublished doctoral dissertation, Michigan State University.

This dissertation deals with the acquisition of reflexive pronouns in second language acquisition. Related to the question of methodology, the topic of this chapter, is a comprehensive discussion of the results from different studies as a function of the methodology adopted.

UNANNOTATED BIBLIOGRAPHY

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language, 72,* 32–68.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In. E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245–261). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bloom, B. (1954). The thought processes of students in discussion. In S. J. French (Ed.), *Accent on teaching: Experiments in general education* (pp. 23–46). New York: Harper.

Corder, S. P. (1973). The elicitation of interlanguage. In J. Svartvik (Ed.), *Errata: Papers in error analysis* (pp. 36–48). Lund, Sweden: CWK Gleerup.

Cowan, R., & Hatasa, Y. (1994). Investigating the validity and reliability of native speaker and second-language learner judgments about sentences. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 287–302). Hillsdale, NJ: Lawrence Erlbaum Associates.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

Eckman, F. (1994). Local and long-distance anaphora in second-language acquisition. In. E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 207 225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ellis, R. (1990). Grammaticality judgments and learner variability. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the Tenth Meeting of the Second Language Research Forum* (pp. 25 60). Eugene. OR: University of Oregon, Department of Linguistics.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition, 13*, 161 186.

Finer, D., & Broselow, E. (1986). Second language acquisition of reflexive binding. *Proceedings of the North Eastern Linguistics Society, 16,* 154 168.

Gass, S. (1994). The reliability of second-language grammaticality judgments. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 303 322). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. (in press). Sentence matching: A reexamination. *Second Language Research*.

Goss, N., Ying-Hua, Z., & Lantolf, J. (1994.). Two heads may be better than one: Mental activity in second-language grammaticality judgments. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 263 286). Hillsdale, NJ: Lawrence Erlbaum Associates.

Han, Y. (2000). Grammaticality judgment tests: How reliable and valid are they? *Applied Language Learning, 11*, 177 204.

Hirakawa, M. (1990). A study of the L2 acquisition of English reflexives. *Second Language Research, 6*, 60 85.

Juffs, A. (this volume). Psycholinguistically oriented second language research.

Lakshmanan, U., & Teranishi, K. (1994). Preferences versus grammaticality judgments: Some methodological issues concerning the governing category parameter in second-language acquisition. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 185 206). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive implicit negative feedback? *Studies in Second Language Acquisition, 22*, 471 497.

Markee, N. (2000). *Conversation analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 227 243). Hillsdale, NJ: Lawrence Erlbaum Associates.

Plough, I., & Gass, S. (1999, March). Measuring grammaticality: A perennial problem. Paper presented at American Association for Applied Linguistics Conference, Stamford, CT.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press

Schachter, J., & Gass, S. (1996). *Second language classroom research: Issues and opportunities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10,* 209 231.

Smagorinsky, P. (this volume). Rethinking protocol analysis from a cultural perspective.

Tarone, E., Gass, S., Cohen, A. (Eds.) (1994). *Research methodology in second-language acquisition* Hillsdale, NJ: Lawrence Erlbaum Associates.

Thomas, M. (1989). The interpretation of English reflexive pronouns by non-native speakers. *Studies in Second Language Acquisition, 11,* 281 303.

White, L., Bruhn-Garavito, J., Kawasaki, T., Pater, J., & Prévost, P. (1997). The researcher gave the subject a test about himself: Problems of ambiguity and preference in the investigation of reflexive binding. *Language Learning, 47*, 145 172.

Yule, G. (1997). *Referential communication tasks*. Mahwah, NJ: Lawrence Erlbaum Associates.

https://doi.org/10.1017/S0267190501000137 Published online by Cambridge University Press