SSLA, **19**, 131–143. Printed in the United States of America.

SECOND LANGUAGE ACQUISITION RESEARCH IN THE LABORATORY

Possibilities and Limitations

Jan H. Hulstijn
Vrije Universiteit, Amsterdam

This paper discusses some possibilities and limitations of laboratory research methods for testing theories of second language acquisition. The paper includes a review of 20 experimental lab studies. The review focuses on the motivation for conducting lab studies, the use of artificial or semiartificial language structures, and various design features (including pre- and posttesting, number of subjects, random subject assignment, between- and within-subjects comparisons, treatment materials and procedures). The paper calls for lab studies addressing issues central to SLA theory ("learning" vs. "acquisition") and ends with some methodological recommendations, concerning the length of experimental treatments, the use of grammaticality judgment tasks, the measurement of reaction times, and the use of retrospective interviews.

The main aim of this special issue of *SSLA* is to explore the possibilities and limitations of laboratory research methods for testing SLA theories as well as to present five lab studies.

It is often difficult to conduct empirical SLA research in the environments where second languages are learned (at home, in the street, at work, and in classrooms) because of the great number of potentially interfering variables in such natural environments. One of the most difficult methodological challenges is to keep all such variables constant. This is almost impossible in "normal" classrooms with real

Address correspondence to Jan Hulstijn, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; e-mail: hulstijn@let.vu.nl.

© 1997 Cambridge University Press 0272-2631/97 $7.50 + .10

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

L2 learners. It comes as no surprise, therefore, that the outcomes of studies conducted in natural learning situations, including classrooms, often form the object of considerable disagreement. This is illustrated by the controversy between Krashen and Lightbown and Pienemann in *TESOL Quarterly* (1993, pp. 717–725) and can be further illustrated with reference to chapter 14 of R. Ellis's *The Study of Second Language Acquisition* (1994). This chapter provides an impressive review of more than 50 empirical studies on formal instruction and SLA. These studies have produced little in the way of hard evidence, however. One of the main reasons for this is that many intervening variables could not be adequately controlled, as Ellis acknowledges in the concluding section of his chapter.

It may therefore be advantageous to abstract away from real-life and classroom learning situations and to conduct empirical research within the boundaries of the research laboratory where intervening variables can be better controlled or manipulated.

There are various ways to bring instruction and learning under the control of the researcher. The first way is to control the language to be learned. This can be done by teaching an artificial or partly artificial language. The advantage of this method is that the researcher can be certain that no subject in the experiment can have advance knowledge of the target structures to be learned and that performance on tests must stem from learning during the experiment.

The second way is to control the nature of the instruction proper (e.g., the explanation of grammar rules). One way of doing this is to replace the live explanation of the teacher by a prerecorded explanation. Using a computer-controlled learning setting is an obvious way of controlling instruction.

Third, the researcher may control the input in a quantitative way by specifying in advance how much instruction and practice subjects will receive. Here again, a computer-controlled learning setting presents itself as an obvious choice.

Finally, the researcher can examine learners' responses during treatment and during testing. Using the computer allows the researcher to measure responses, as well as reaction times as an indication of automaticity.

Lab research, however, has its limitations and practical disadvantages. Because such research deliberately abstracts away from real-life learning situations, it simultaneously limits the possibilities to extrapolate their findings legitimately to real-life learning. Furthermore, as Schmidt (1994b, pp. 166–167) observes, in some laboratory studies subjects are instructed to do things that are not likely to be reflected in present-day language pedagogy (e.g., memorizing examples). Therefore, without additional research in real L2 learning environments, one should be extremely cautious in drawing immediate conclusions from laboratory studies to language pedagogy.

REVIEW

To date, relatively little L2 learning or acquisition research has been conducted in lab situations, but in recent years we have seen some innovative studies.[1] There is room for more SLA laboratory research, and one of the functions of this issue of *SSLA* is to give examples of SLA lab research.

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

Although the focus of this review will be on lab methods, it will be stressed right from the outset that good empirical research must be based on theoretically well-motivated questions and that the adequacy of research methods can only be assessed from the perspective of the theoretical questions addressed. All of the following comments should be interpreted with this preliminary proviso in mind.

Included in this review are 20 studies, published since 1988, that meet the following criteria:

1. They addressed issues pertinent to theories of SLA. Therefore, applied studies, such as investigations into computer-aided language learning (e.g., Nagata & Swisher, 1995), and nonlinguistic studies, such as the learning of meaningless strings of symbols (e.g., studies on so-called Miniature Artificial Languages as conducted by Reber and his associates [Reber, 1989], were excluded).
2. They all comprised a learning phase of some sort. Excluded were, therefore, studies only consisting of a testing phase, such as studies in which computerized tests were administered to various groups of subjects, without experimentally controlling previous input exposure or instruction.
3. In all studies input exposure and/or instruction were experimentally controlled and/or manipulated, replacing a live teacher by either a computer or a human experimenter. Learner–teacher interaction was under complete control in that the computer or experimenter reacted on learner responses in a completely consistent, prescribed way.
4. They involved some sort of grammar learning (morphophonology and syntax). Excluded are studies that exclusively investigated pronunciation and spelling (e.g., Michas & Berry, 1994), as well as the extensive experimental literature on vocabulary learning.
5. All studies have been published in accessible journals or books (with ISBN or ISSN numbers).

The 20 studies included in this review are, in alphabetical order, Alanen (1995), Carroll and Swain (1993), Carroll, Swain, and Roberge (1992), Cook (1988), de Graaff (this issue), DeKeyser (1995, this issue), Doughty (1991), N. Ellis (1993), N. C. Ellis and Schmidt (this issue), Hulstijn (1989c), Issidorides (1988), Leow (1993), Robinson (1996, this issue), Robinson and Ha (1993), Shook (1994), VanPatten (1990), Yang and Givón (this issue), and Zekhnini and Hulstijn (1995). There are probably more studies that meet the preceding criteria, but these could not be traced before the completion of this text (summer 1996).

Theoretical Issues

1. Most studies address one or both of the following two issues (the terminology is that of Schmidt, 1994a, p. 20):

(a) Explicit versus implicit learning, that is, learning with or without an awareness of learning at the point of learning.
(b) Explicit instruction versus input enhancement, that is, the provision of explicit rules versus making input especially salient and likely to be attended.

Theoretical claims that these studies try to support or falsify are those of Krashen (1981, 1982), Reber (1989), Long (1988), N. Ellis (1993), Schmidt (1994a, 1994b), Carr

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

and Curran (1994), and many others. Reviewed studies addressing either or each issue are by Doughty (1991), N. Ellis (1993), DeKeyser (1995), Hulstijn (1989c), Robinson (1996, this issue), and VanPatten (1990). In some of these studies, the explicit–implicit factor is crossed with a linguistic factor (e.g., simple versus complex rules, purely formal versus form-meaning rules, categorical versus prototypical rules), as in Alanen (1995), de Graaff (this issue), DeKeyser (1995), and Robinson (1996).

2. Some studies address the issue of rule versus item learning (de Graaff, this issue; DeKeyser, 1995; Ellis & Schmidt, this issue; Robinson, 1996, this issue; Zekhnini & Hulstijn, 1995). To what extent should language acquisition be seen as the learning of abstract rules and principles (Chomsky, 1986; Pinker, 1989) or rather as the acquisition of specific exemplars, the acquisition of similarities between exemplars (cues), and the generalization of cues (the associationist, connectionist perspective, represented by MacWhinney [1989])?

3. Some studies (de Graaff, this issue; DeKeyser, this issue; Robinson, this issue; Robinson & Ha, 1993; Yang & Givón, this issue) address the issue of skill acquisition: Is automaticity in using language achieved through a gradual change from controlled processing on the basis of declarative knowledge to automatic processing on the basis of procedural knowledge (Anderson, 1983) or through a mechanism of strengthening a single-step, direct retrieval of past episodes stored in memory (e.g., Logan, 1988)?

4. Four studies (Carroll & Swain, 1993; Carroll et al., 1992; Cook, 1988; Doughty, 1991) address the issue of "learnability," as it has become known in the linguistic literature (Gregg, 1996; Pinker, 1989). How can the acquisition and induction of abstract principles of syntax be explained when exposure to input (a) is limited and/or (b) does not contain negative information? To what extent is negative feedback (error correction) required to restrict the domain of overgeneralizations? Chomsky's theory of Universal Grammar and various implicational theories of markedness, taking into account differences between the grammars of learners' L1 and L2 (e.g., Eckman, 1996; Hawkins, 1987) give contrasting views and need to be put to the test.

5. Three studies (Issidorides, 1988; Leow, 1993; Yang & Givón, this issue) address the issue of L2 learning under conditions of simplified input.

6. Some studies address the issue of focus of attention and input enhancement during input processing (Alanen, 1995; Hulstijn, 1989c; Leow, 1993; Shook, 1994; VanPatten, 1990).

Motivation for a Laboratory Study

All investigators of the studies reviewed here chose for a lab study in order to have complete control of exposure to input and instruction, thereby eliminating potentially confounding variables present under normal language learning conditions, in or outside teacher-guided language courses. As Carroll and Swain put it, "If feedback does not work in an experimental situation, it is highly unlikely that it would work elsewhere" (1993, pp. 361–362).

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

Use of Artificial or Semiartificial Target Language and Target Structures

Nine studies used English or Spanish as the target language with learners of English or Spanish as a second or foreign language as subjects (Carroll & Swain, 1993; Carroll et al., 1992; Doughty, 1991; Leow, 1993; Robinson, 1996, this issue; Robinson & Ha, 1993; Shook, 1994; VanPatten, 1990). Most of these studies therefore had to include pretests through which individuals already familiar with the target structures could be excluded from the analyses. The use of a natural language as the target language and L2 learners of that language as subjects bears a well-known inherent threat to the internal validity of the study, namely, that not all subjects included in the study are exactly the same in terms of prior knowledge. This fact may be more harmful in one study than in another, depending on the question researched.

Three studies used a natural (or slightly modified) language as input to learners who, with certainty, had had no prior exposure to these languages: N. Ellis (1993) with Welsh as the target language and English subjects, Issidorides' (1988) second experiment with Dutch as the target language and Greek subjects in Greece, and Alanen (1995) with (modified) Finnish as the target language and English subjects as subjects.

Six studies used self-constructed artificial languages, whose sentences all bore referential meaning (all input sentences referring to possible states of affairs in the extralinguistic world) and whose grammar rules all remained within the grammatical constraints of the world's natural languages (Cook, 1988; de Graaff, this issue; DeKeyser, 1995, this issue; Ellis & Schmidt, this issue; Yang & Givón, this issue). Using an artificial language allows, as Cook (1988, p. 509) put it, for complete control of the consistency and purity of input. In the studies of DeKeyser (1995, this issue), the meanings of the relevant lexical and grammatical morphemes were chosen such that they could easily be presented pictorially, a useful technique for lab studies.

Four studies used partly natural and partly artificial input. Alanen (1995) deliberately omitted a number of features from the target language (Finnish) in her study in order to make the learning task less difficult and the target structures more salient. Hulstijn (1989c) confronted Dutch subjects with sentences made up of Dutch content words (so that the meaning of the sentences could be easily understood) but containing some artificial morphemes and an artificial word order (the learning targets). Zekhnini and Hulstijn (1995) investigated how native and nonnative speakers of Dutch learned article–noun pairs, consisting of either one of the two Dutch definite articles (*de* and *het*) followed by a pseudo-Dutch noun. The aim of the study was to falsify the claim that learners obligatorily encode noun gender as an inherent noun feature. Robinson (this issue) investigated the acquisition of English dative alternation using artificial (pseudo-)verbs (divided into a subclass that does and a subclass that does not allow dative alternation), in otherwise natural sentences of English. By doing so, he was sure that, at the beginning of the treatment, subjects did not know the verbs to be taught. (This did not rule out the possibility, however, that subjects already knew the dative alternation rule.)

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

Some studies, notably the ones by Hulstijn and his associates (de Graaff, this issue; Hulstijn, 1989b, 1989c; Issidorides, 1988; Zekhnini & Hulstijn, 1995), adopted a so-called twin approach (see also Hulstijn & de Graaff, 1994). This means that an experiment using completely artificial or partly artificial input was paralleled by a similar experiment using a natural language. The experiment with (partly) artificial input ranks relatively high on reliability (complete control of L2 learners' prior knowledge) but possibly low on (ecological) validity. This is offset, however, by the accompanying experiment with natural language input ("real" L2 learners learning a "real" L2), ranking relatively high on ecological validity but possibly low on reliability. The researcher then hopes that the results of the twin experiments dovetail nicely, allowing for interpretations that can be credited with both reliability and validity.

Relevant Design Features

Pretesting. All studies using natural language input, except Robinson (1996) and Robinson and Ha (1993), included pretests in order to select subjects with desired levels of (un)familiarity with the target structures or take account of subjects' prior knowledge in other ways (e.g., via pretest–posttest covariance analyses). N. Ellis (1993) administered a language learning aptitude test and used its results in assigning subjects to experimental conditions.

Number of Subjects. Two studies (Alanen, 1995; Doughty, 1991) used fewer than 10 subjects per experimental group. Use of more subjects might have increased these studies' reliabilities. Most studies had 20–30 subjects per experimental group.

Random Subject Assignment. In some studies, subjects were randomly assigned to conditions (de Graaff, this issue; Doughty, 1991; N. Ellis, 1993; Robinson, 1996). In four studies (Hulstijn, 1989c; Leow, 1993; Shook, 1994; VanPatten, 1990), the activities in the learning and testing phases were group-administered, and groups rather than individual subjects were randomly assigned to conditions. In some studies, subjects were matched, for example, for aptitude (N. Ellis, 1993) or for sex (Alanen, 1995; Yang & Givón, this issue) and former language learning experience (Alanen, 1995; DeKeyser, 1995, this issue).

Three studies (Carroll & Swain, 1993; Carroll et al., 1992; Cook, 1988) do not give explicit information whether subjects were randomly assigned to experimental conditions.

Between- and Within-Subjects Comparisons. Most studies adopted a between-subject design to compare different input exposure and instruction regimes. In some studies, subject groups performed the same tasks but were exposed to different types or amounts of language input (Cook, 1988; Issidorides, 1988; Robinson, this issue). In many studies, subject groups were exposed to the same verbal input but received different information about the input—for example, more or less explicit

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

grammar explanation or different feedback regimes (Alanen, 1995; Carroll & Swain, 1993; Carroll et al., 1992; de Graaff, this issue; Doughty, 1991; N. Ellis, 1993; Robinson, 1996)—or they were required to perform different tasks with the input (DeKeyser, this issue; Hulstijn, 1989c; VanPatten, 1990).

Two studies (Ellis & Schmidt, this issue; Robinson & Ha, 1993) adopted an exclusively within-subjects design; frequency of exposure to targets was implemented as a within-subjects factor.

Many studies crossed a between-subjects instructional factor with a within-subjects linguistic factor, such as simple versus complex rules (de Graaff, this issue; Robinson, 1996), categorical versus prototypical rules (DeKeyser, 1995), lexical versus nonlexical rules or targets (de Graaff, this issue; Yang & Givón, this issue), and purely formal versus form-meaning rules (Alanen, 1995).

Control Groups. Four studies included a control group, which was exposed to the input language without further information or feedback (Alanen, 1995; Carroll & Swain, 1993; Carroll et al., 1992; N. Ellis, 1993). There is only one study that included a no-treatment control group, that is, a group that was not exposed to the target language but only performed the pre- and posttests (Experiment 2 in Hulstijn, 1989c). The inclusion of this group turned out to be very revealing, in that subjects in this group exhibited an increase in performance from pretest to posttest although they had not been exposed to the targets in-between. The interval between pre- and posttest, however, was only 30 minutes, which might have caused a retest effect.

Computer Administration. Most of the more recent studies used the computer for input presentation, learning instructions, feedback (if applicable), and the elicitation and registration of responses, with or without reaction times. Two studies used a random-access tape-slide projector with a built-in cassette player and microprocessor, for a synchronized presentation of visuals and sound to individual subjects (Issidorides, 1988) or to small groups of subjects (Zekhnini & Hulstijn, 1995).

Information Concerning Treatment and Testing Procedures. Some studies give too little information concerning treatment matters to allow for a proper assessment of the study's validity. Sometimes it is not exactly clear what the learning task was and what sort of instructions were given to subjects. A researcher labeling a treatment condition as eliciting *implicit* or *incidental* learning processes has the onus of describing materials, task, performance conditions, and instructions in sufficient detail so as to make these labels credible (see Hulstijn, 1989a). Studies lacking sufficient detail in this respect are Alanen (1995) and Carroll and Swain (1993).[2]

Retrospective interviews or questionnaires, administered after treatment and testing, may reveal how subjects had experienced the treatment tasks and had interpreted task instructions. In one study (Alanen, 1995), subjects had to think aloud during treatment. This procedure may indeed give the required information but, given its metacognitive nature, has the drawback of interfering with the intended way of input processing in the case of subjects placed in conditions of implicit or incidental learning.

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

***Length of Treatment* (*Exposure, Instruction, and Practice*).** The duration of the learning phase was short in most studies, treatment being limited to a period of between a few minutes and approximately 1 hour. Learning periods this short may form a threat to the external validity of the study (depending, of course, on the question investigated). Only five studies involved longer periods of exposure and learning (de Graaff, this issue; DeKeyser, 1995, this issue; Doughty, 1991; Yang & Givón, this issue). Four of these studies allowed for a within-subjects, longitudinal investigation of skill acquisition (de Graaff, this issue; DeKeyser, 1995, this issue; Yang & Givón, this issue).

The sittings for subjects may have been quite boring to subjects in some studies. This seems especially true in the N. Ellis (1993) study, because subjects in that study spent 3–8 hours (an average of almost 1,400 trials) working on a single, purely formal, nonsemantic phonological and ortographical rule, involving only some dozens of words and sentences. Given the small number of words and sentences involved (no exact figures are given), this study might be considered a borderline case of genuine language learning and purely cognitive concept formation. This point will be taken up below.

Posttesting. Many studies included only one posttest task, measuring performance in one modality or in one format only. This may also be a threat to the validity of the study.

One study (Hulstijn, 1989c) justifies the selection of the posttest tasks in terms of the concept of transfer appropriate processing (Bransford, Franks, Morris, & Stein, 1979) and attempts to take into account differences in compatibility between the task(s) performed in the treatment phase and task(s) performed in the testing phase.

Many studies used a grammaticality judgment task of one type or another. Some used such a task as the only measure of learning (Carroll & Swain, 1993; Cook, 1988; Robinson, 1996; Robinson & Ha, 1993). In the SLA literature, concerns have been expressed concerning the validity of grammaticality judgments (Beck, 1992; Birdsong, 1994; Bley-Vroman & Masterson, in press; Gass, 1994; Sorace, 1996), especially when elicited from nonnative speakers exclusively (precluding a comparison with native speaker performance). Furthermore, many studies measured the reaction times of subjects' responses (again mostly grammaticality judgments). The use of speeded judgment tasks (de Graaff, this issue; DeKeyser, 1995; N. Ellis, 1993; Yang & Givón, this issue) raises the question of when such tests can be said to tap implicit knowledge and when explicit knowledge. Reaction times of nonnative speakers, in comparison to those of native speakers, are often slow or show large variability (standard deviations), as in the study of Robinson and Ha (1993). In such cases, reaction time data should be interpreted with much caution. Of course, one cannot simply draw a dividing line between slow and fast reaction times and interpret slow and fast reaction times as evidence of nonautomatic and automatic processing, respectively. Thus, one cannot extrapolate reaction times findings to implicit knowledge. This may limit the usefulness of reaction-time measurements in nonlongitudinal studies. Longitudinal studies, however, allow the researcher to investigate whether and how reaction times decrease over time. They thus allow the researcher to

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

witness the gradual emergence of automatization. All studies in this issue but one registered reaction times longitudinally.

An elegant feature in the Yang and Givón study (this issue) is that subjects were not only tested on the artificial language that they had learned (with reaction times as one dependent variable) but also on their native language, English, allowing for an assessment of the validity of the artificial language measures. This study also involved the administration of not just one but three types of word recognition tasks, allowing for a better assessment of the validity of the data of each of these tasks.

Some studies (e.g., Carroll & Swain, 1993) lack information concerning the way in which subjects' responses were scored and coded for analysis.

Two studies used a sentence repetition test, either in written form, with 10 seconds presentation time (Hulstijn, 1989c), or in oral form (Yang & Givón, this issue). The use of sentence repetition tests was not successful in the former study, because even subjects in the no-treatment control group were able to write down a substantial number of the artificial target features.[3]

Retrospective Measures. Two studies (de Graaff, this issue; DeKeyser, 1995) with learning phases of several weeks were rounded off with individual oral interview sessions or the filling out of questionnaires in order to assess to what extent subjects had actually acquired explicit knowledge of the target rules and to what extent they had performed the assigned treatment tasks as intended by the researchers.

CONCERNS AND RECOMMENDATIONS

Laboratory research offers a number of important advantages over research conducted with L2 learners in classrooms or with uninstructed, so-called natural learners: control of the language and the target structures to be learned, control of exposure, control of instruction (explanation), control of tasks, and control of response measurement. It is therefore mandatory to further pursue this promising research method, along with the more traditional classroom and natural research methods. In its zeal to attain full control of learning and instruction, however, laboratory research may run the risk of losing sight of some issues, to be mentioned in this closing section. The concerns raised and recommendations made in this section might be appropriate, depending on the research questions investigated. With this proviso in mind, researchers should be encouraged to critically examine the following points.

1. One might wonder whether some of the lab studies reviewed here (e.g., the ones by N. Ellis) investigated purely cognitive concept formation rather than (second) language learning. Most SLA researchers will argue that language learning is different from the learning of nonlinguistic symbols, the linear order of appearance of these symbols, and their (sub)category membership (exhibited by their morphonological form). N. Ellis (1996) and Ellis and Schmidt (this issue), however, openly challenge linguists to demonstrate that there are rules of grammar that cannot be learned on the basis of associative processes. This calls for lab research not limited to the

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

acquisition of morphological subcategorization rules (e.g., past-tense formation of verbs, pluralization of nouns) nor to the acquisition of formal dependencies, whether local or nonlocal (as in the second experiment, reported by Ellis & Schmidt, this issue). Lab studies should also address the logical problem of (second) language learning (Gregg, 1996). Psychologists and linguists should be encouraged to join their expertise in designing lab studies focusing on the question of whether and to what extent L2 learning should be distinguished from L2 acquisition, as in (a) and (b), respectively (see Schwartz, 1993, and Zobl, 1995, for testable redefinitions of the notions learning and acquisition; see Pinker, 1991, for a distinction between rules and principles):

(a) To what extent is second language learning a gradual process of skill acquisition? Which linguistic principles, rules, or elements must be learned on the basis of principles of association? Consequently, to what extent is language learning dependent on input frequency and practice, and to what extent does it benefit from explicit instruction?
(b) To what extent must second language learning be seen as the acquisition of abstract linguistic knowledge (in the sense of Chomsky's competence)? Which principles and parameter values belonging to the core of the grammar of the target language, and which peripheral rules or elements must be acquired in a largely subconscious way similar to the way in which children acquire the grammar of their native language? Consequently, to what extent are explicit instruction, error correction, input frequency, and practice irrelevant for second language learning?

The studies of Yang and Givón, DeKeyser, and de Graaff (all three included in this issue) show that it is possible to motivate individuals to devote themselves to the learning of an artificial language during several weeks in a setting resembling "real" second or foreign language instruction. This is an important methodological feat and should encourage scholars interested in the logical problem of (second) language learning (e.g., the resetting of parameters, the acquisition of clustered parameter properties under exposure to only one property, the acquisition of a rule under exposure to different positions in accessibility hierarchies) to devise experiments adopting a similar method.

2. Many of the studies reviewed here exposed subjects to L2 input for extremely short periods or to extremely few stimuli. This may form a threat to their validity. Some issues of SLA, such as claims concerning the impossibility or inferiority of implicit or unattentional learning, can be studied in a valid way only with L2 exposure longer than just a single session of 15–60 minutes and only with input containing large numbers of relevant instances in the input.

3. There is room for concern about the use of grammaticality judgments with nonnative speakers. In general, it is preferable to measure a construct (e.g., acquired knowledge of a target structure) with more than one task or under more than one task condition.

4. There is room for concern about the measurement of reaction times of responses given by nonnative speakers in nonlongitudinal investigations, especially when reaction times are relatively slow and show large variability. In general, it is

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

preferable to compare subjects' reaction times to L2 stimuli with reaction times to comparable L1 stimuli.

5. In general, it is preferable to end the investigation with retrospective interviews or questionnaires in order to assess whether subjects had experienced and performed their treatment tasks as intended by the investigators and to measure any explicit, verbalizable knowledge of the target rules.

NOTES

1. No distinction in this paper is made between acquisition and learning unless explicitly specified.
2. This is a criticism of the reports rather than of the investigations themselves.
3. See Bley-Vroman and Chaudron (1994) for a critical appraisal of the use of elicited imitation as a measure of L2 competence.

REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Honolulu: University of Hawai'i Press.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.

Beck, M.-L. (1992, April). *Grammaticality judgments: What subjects could have told us if we'd only bothered to ask*. Paper presented at the 1992 Second Language Research Forum, Michigan State University, East Lansing.

Birdsong, D. (1994). Asymmetrical knowledge of ungrammaticality in SLA theory. *Studies in Second Language Acquisition, 16,* 463–473.

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

Bley-Vroman, R., & Masterson, D. (in press). Reaction time as a supplement to grammaticality judgements in the investigation of second language learners' competence. *Second Language Research*.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331–354). Hillsdale, NJ: Erlbaum.

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences: Applications or syntax. *Studies in Second Language Acquisition, 16,* 205–230.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15,* 357–386.

Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics, 13,* 173–198.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.

Cook, V. J. (1988). Language learners' extrapolation of word order in micro-artificial languages. *Language Learning, 38,* 497–529.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19,* 249–276.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379–410.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19,* 195–222.

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13,* 431–469.

Eckman, F. (1996). A functional–typological approach to second language acquisition theory. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 195–211). San Diego: Academic Press.

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology, 5,* 289–318.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18,* 91–126.

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition, 19*, 145–172.
Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
Gass, S. M. (1994). The reliability of second-language grammaticality judgments. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 303–322). Hillsdale, NJ: Erlbaum.
Gregg, K. (1996). The logical and developmental problems of second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 50–81). San Diego: Academic Press.
Hawkins, J. A. (1987). Implicational universals as predictors of language acquisition. *Linguistics, 25,* 453–473.
Hulstijn, J. H. (1989a). A cognitive view on interlanguage variability. In M. R. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 17–31). New York: Plenum Press.
Hulstijn, J. H. (1989b). Experiments with semi-artificial input in second language acquisition research. In B. Hammarberg (Ed.), *Language learning and learner language*. Papers from a conference held in Stockholm and Åbo, October 17–18, 1988. *Scandinavian Working Papers on Bilingualism, 8,* 28–40 (issued from the Centre for Research on Bilingualism, University of Stockholm).
Hulstijn, J. H. (1989c). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H. W. Dechert & M. Raupach (Eds.), *Interlingual processing* (pp. 49–73). Tübingen: Gunter Narr.
Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11,* 97–112.
Issidorides, D. C. (1988). The discovery of a miniature linguistic system: Function words and comprehension of an unfamiliar language. *Journal of Psycholinguistic Research, 17,* 317–339.
Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
Krashen, S. (1993). The effect of formal grammar teaching: Still peripheral. *TESOL Quarterly, 27,* 722–725.
Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition, 15,* 333–355.
Lightbown, P. M., & Pienemann, M. (1993). Comments on Stephen D. Krashen's "Teaching issues: Formal grammar instruction." *TESOL Quarterly, 27,* 717–722.
Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review, 95,* 492–527.
Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115–142). Rowley, MA: Newbury House.
MacWhinney, B. (1989). Competition and connectionism. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 422–457). Cambridge: Cambridge University Press.
Michas, I. C., & Berry, D. C. (1994). Implicit and explicit processes in a second-language learning task. *The European Journal of Cognitive Psychology, 6,* 357–381.
Nagata, N., & Swisher, V. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals, 28,* 337–347.
Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.
Pinker, S. (1991, August 2). Rules of language. *Science*, pp. 530–535.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.
Robinson, P. J. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18,* 27–67.
Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–248.
Robinson, P. J., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition, 15,* 413–438.
Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11,* 11–26.
Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London: Academic Press.
Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15,* 147–164.
Shook, D. J. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning, 5,* 57–93.
Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–409). San Diego: Academic Press.

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12,* 287–301.

Yang, L. R., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition. *Studies in Second Language Acquisition, 19*, 173–194.

Zekhnini, A., & Hulstijn, J. H. (1995). An experimental study on the learning of arbitrary and non-arbitrary gender of pseudo Dutch nouns by nonnative and native speakers of Dutch. *Toegepaste Taalwetenschap in Artikelen* (the journal of Anéla, the Dutch affiliate of AILA), *53,* 121–136.

Zobl, H. (1995). Converging evidence for the 'acquisition-learning' distinction. *Applied Linguistics, 16,* 35–56.

https://doi.org/10.1017/S0272263197002015 Published online by Cambridge University Press