SSLA, **19**, 195–221. Printed in the United States of America.

BEYOND EXPLICIT RULE LEARNING

Automatizing Second Language Morphosyntax

Robert M. DeKeyser
University of Pittsburgh

This study is a fine-grained analysis of extensive empirical data on the automatization of explicitly learned rules of morphosyntax in a second language. Sixty-one subjects were taught four morphosyntactic rules and 32 vocabulary items in an artificial language. After they had reached criterion on a set of metalinguistic tests of grammar and vocabulary, they engaged in systematic, computer-controlled comprehension and production practice for 8 weeks. Comprehension practice consisted of choosing between pictures displayed on the computer screen to match a sentence; production practice consisted of typing the correct sentence corresponding to a picture. All subjects were taught the same rules and then practiced them, and all subjects had the same amount of comprehension and production practice, but which rules were practiced in comprehension and which in production varied between groups. Results show that the learning of morphosyn-

This study was supported by U.S. Department of Education grant P017A50064. I thank Chris Connors and Jim Rankin for their expert programming and Ben and Philippa Benson-Xu, Jeanine Carlock, Lorien Clemens, Jing-Fu Fan, Kiduk Kim, Jannine Markizon, Jeri and Scott Misler, Yong-Ping Mou, David Novinksi, John Smith, David Steinitz, and Zander Teller for their superb acting performances. Thanks are also due to the experimenters Keiko Iijima, Jannine Markizon, Andre Mather, Don Peckham, Leonore Rodrigues, Michelle Sadlier, Clay Taylor, Paul Toth, Eugenia Wan, and Bill Williams. I gratefully acknowledge the advice and encouragement from Carol Baker, Alan Juffs, Donald McBurney, Daniel Everett, Robert Henderson, David Malicki, Christina Paulston, Charles Perfetti, and Richmond and Sarah Thomason and the financial support of the University of Pittsburgh Central Research Development Fund for a pilot study as well as that of the Linguistics Department for various expenses. Kathleen Bardovi-Harlig, Nick Ellis, Jan Hulstijn, Donald Peckham, Peter Robinson, and Lynne Yang provided helpful comments on an earlier draft of this paper. Finally, I wish to express my utmost gratitude to Jim Rankin for computational feats far beyond the call of duty.

Address correspondence to Robert DeKeyser, Department of Linguistics, University of Pittsburgh, Pittsburgh, PA 15260; e-mail: rdk1@vms.cis.pitt.edu.

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

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

tactic rules is highly skill-specific and that these skills develop very gradually over time, following the same power function learning curve as the acquisition of other cognitive skills. These results are consistent with current skill acquisition theory.

In recent years a number of empirical studies have been forthcoming that show the superiority of explicit over implicit processes in second language learning, at least for certain rules and certain learners (e.g., Alanen, 1995; de Graaff, this issue; DeKeyser, 1995; N. Ellis, 1993; House, 1996; Master, 1994; Robinson, 1996). A weakness of these studies is that they stop short of demonstrating that the explicitly learned knowledge can be used automatically, that is, fast, with a low error rate, and with little interference from other cognitive tasks.

Whereas a substantial literature on automaticity and automatization exists in cognitive psychology, and the automatization of rules has long been recognized as a central problem in applied linguistics by those who believe rules are useful at all, hardly any fine-grained experimental literature is available on the process of rule automatization in the course of second language learning. Robinson and Ha (1993) provided evidence for a certain degree of automaticity in grammaticality judgments for sentences in a pseudo-artificial language, but they did not investigate the development of automaticity over time. The present study aims to contribute to filling the gap in empirical research on automatization by presenting a fine-grained longitudinal account of the acquisition of the rules of an artificial language (a miniature linguistic system) under different conditions of practice.

AUTOMATICITY

Automaticity has received considerable attention in the cognitive literature of the last 20 years. Despite continuing disagreements about the exact nature of capacity limitations and the role of attention (cf., e.g., Pashler, 1994) about the relationship between automaticity and implicit knowledge (Seger, 1994), and about how automaticity is brought about (with competing models as a result, such as Anderson's [1993] ACT-R (Adaptive Control of Thought—Rational) model; Schneider and Detweiler's [1988] connectionist-control model, and Logan's instance learning model [Logan, 1988, 1992; Logan & Etherton, 1994; Logan, Taylor, & Etherton, 1996]), substantial progress has been made. It is now clear that there is a continuum of automaticity rather than an automatic-controlled dichotomy. Researchers widely agree that phenomena such as gradual drop-offs in reaction time and error rates, and diminished interference from and with simultaneous tasks (cf., e.g., Logan & Etherton, 1994), are reliable criteria of automaticity. Some of the experiments showing these effects have become classics (e.g., Schneider & Shiffrin, 1977).

The most widely accepted theory on how automaticity is brought about is Anderson's ACT[1] model of the human cognitive architecture (cf., esp. Anderson, 1987, 1992, 1993). According to this model, knowledge typically starts out as explicit

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

(declarative) information, "knowledge that," which is turned into specialized procedural rules, and "knowledge how," for very specific behaviors through analogy with a series of examples and with the help of very general behavioral rules. It is then fine-tuned over time as a function of cost-effectiveness (probability of being correct and cost in terms of mental resources). The result of this last process is a gradual drop-off in reaction time and error rate. This drop-off is generally held to take the form of a power function (Newell & Rosenbloom, 1981; Anderson, 1993), that is, a function of the form $y = x^n$, reflecting that the amount of improvement (the change in y, be it reaction time or error rate) decreases as a function of the amount of practice x, the power n being negative.

The ACT model has been tested very extensively and successfully in empirical studies on the learning of algebra, geometry, and computer programming. Anderson's (1983) application of ACT* to first language learning, however, has remained controversial, because of its insistence (subsequently relaxed; see Anderson & Fincham, 1994) that all knowledge starts out in declarative form.

In what follows, I will draw on rule-based theories of automaticity in general and on Anderson's ACT-R model in particular. For an empirical study drawing on memory-based theories of automaticity in general, and Logan's instance theory in particular, see Robinson (this issue).

AUTOMATIZATION OF SECOND LANGUAGE RULES

Automatization is the process that leads to automaticity. In the field of applied linguistics, the notion of automatization was used loosely by proponents of audiolingualism (e.g., Rivers, 1964) and cognitive code (e.g., Chastain, 1971) and elaborated didactically by early communicative methodologists (Paulston & Bruder, 1976). These authors claimed that automaticity was to be achieved by overlearning of stimulus–reaction chains (Rivers), grammar teaching followed by extensive drill practice (Chastain), and the careful sequencing of mechanical, meaningful, and communicative drills (Paulston & Bruder). At that time, the computational tools for providing strictly controlled treatments and conducting fine-grained data collection and analysis were not available to applied linguists, and no empirical studies were carried out on the exact nature of automatization.

Since the late 1970s, the notion of automatization in second language learning has been questioned by numerous authors. Some, most notably Krashen (e.g., 1982, 1985, 1994), question the usefulness of any rule learning or focus on form and argue that the spontaneous use of a rule after it was taught and practiced by no means results from automatizing that rule but is the result of a completely independent, implicit process of acquisition, resulting from exposure to large amounts of comprehensible, meaningful input. Others, whereas recognizing that focus on form is useful, still see it as a way of improving the implicit intake of structures from the input and therefore consider systematic and form-focused practice in production to be of no use to acquisition (R. Ellis, 1993; VanPatten & Cadierno, 1993a, 1993b).

A variety of recent studies, in both the classroom (e.g., Scott, 1989; Spada & Lightbown, 1993; White, Spada, Lightbown, & Ranta, 1991) and the laboratory (e.g.,

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

Alanen, 1995; Doughty, 1991; N. Ellis, 1993), have provided evidence that some sort of focus on form is useful, and some theoreticians have endorsed the notion of automatizing explicitly learned second language rules through production practice, such as Gatbonton and Segalowitz (1988), Johnson (1996), O'Malley and Chamot (1990), and Schmidt (1992). Johnson (1996), in particular, discusses the conditions for automatization at length, defining "the skill of automatization" (*automaticity* in our terminology) as "the ability to get things right when no attention is available for getting them right" (p. 137). He argues that learners should be given a series of tasks at the level of *ra* – 1 (required attention minus one), that is, where the amount of attention that can be paid to the forms in question is slightly less than comfortable for the learner at each stage. This has traditionally been accomplished by increasing focus on meaning, but Johnson (1996) insists that other kinds of form defocus, such as time pressure and competing tasks are equally useful distraction techniques, and hold promise in the computer age (p. 176).

Virtually no empirical evidence for the usefulness of practice in production exists yet, and certainly no precise documentation of the process of automatization of rules through such practice. The few studies in the domain of second language learning that have built on Schneider and Shiffrin's (1977) methodology have been limited to vocabulary learning (e.g., Favreau & Segalowitz, 1983), reading (e.g., McLeod & McLaughlin, 1986), or word recognition in reading (e.g., Segalowitz & Segalowitz, 1993); see McLaughlin (1987) for a review. On the other hand, the literature on temporal variables (e.g., Möhle, 1984; Raupach, 1987; Towell, Hawkins, & Bazergui, 1996; for a critical review, cf. Griffiths, 1991) provides detailed evidence of speed-up and chunking in advanced learners, but its failure to trace the learning of specific rules prevents it from being convincing evidence for automatization processes. However, several studies, in particular those by Towell et al. (1996) and Raupach (1987), are interesting because of their wealth of fine-grained performance data from different stages of the learning process, even if their database is small, and the interpretation in terms of skill acquisition models is highly speculative.

In summary, the issue of automatization of grammar rules through practice in production remains largely uninvestigated because the applied linguistics research methodology of the 1960s and 1970s was insufficiently developed to address the complexities of automatization in language learning and because second language researchers in the 1980s were reluctant to conduct empirical studies on it, due to the prevailing doubts about the usefulness of learning and practicing rules. But the time is now ripe to carry out such research, given the renewed confidence in teaching and learning at least some kinds of rules (see de Graaff, this issue; DeKeyser, in press; and Hulstijn & de Graaff, 1994, for more discussion of the explicit learning of different kinds of rules) and the experimental and analytical tools now at our disposal for the study of automatization.

THE SPECIFICITY OF PRODUCTION RULES

Related to the controversy over whether automatization exists in second language learning, or whether second languages are learned in a completely different way, due to the autonomous nature of the linguistic module of the mind, is the issue of

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

the specificity of knowledge. Those who claim that linguistic competence in a second language is acquired implicitly also argue that the same knowledge, once acquired, is drawn on for both production and comprehension, making practice in production unnecessary (Krashen, 1985; VanPatten & Cadierno, 1993a, 1993b). Those who argue that all cognitive skills are acquired through the same mechanisms of proceduralization of declarative knowledge see proceduralized knowledge (production rules) as highly specific, so that comprehension and production, for instance, draw on completely different sets of rules and therefore require separate practice (see, e.g., Anderson, 1993, pp. 37–38, where this is illustrated in detail for reading and writing computer code).

Until recently, the issue of skill specificity had received virtually no attention in empirical second language acquisition research. Recent experiments conducted by VanPatten and his associates suggest that practice in production does not contribute anything to acquisition beyond form-focused teaching followed by comprehension practice (Cadierno, 1995; VanPatten & Cadierno, 1993a, 1993b), or at least that comprehension practice alone is enough to bring about significant improvement in both comprehension and production skills (VanPatten & Sanz, 1995). These experiments dealt with the form and position of object clitics (VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz, 1995) and the morphology of the past tense (Cadierno, 1995) in Spanish. DeKeyser and Sokalski (1996), however, have challenged the generalizability of these findings to other elements of morphosyntax (such as the conditional) and have argued that the effect of input and output practice is largely skill-specific.

In all these studies, the period of training was short, and the outcome measures were largely paper-and-pencil tests; even in the case where oral testing was conducted (VanPatten & Sanz, 1995), it allowed for a certain amount of reflection. No research so far has investigated the longitudinal effect of systematic practice on real-time performance in comprehension and production. Therefore, it was decided to conduct a longitudinal study on the differential effect of comprehension practice and production practice in the highly controlled environment of the computer-assisted learning of a miniature linguistic system.

METHOD

Hypotheses

1. Once second language grammar rules have been explicitly assimilated, practice will lead to gradual automatization, as measured by reduced reaction time, reduced error rate, and decreased interference from simultaneous tasks (cf., e.g., Johnson, 1996; Schmidt, 1992).
2. The reduction in reaction time and error rate will follow a power function for individual grammar rules (cf., e.g., Newell & Rosenbloom, 1981).

3a. The practice effect will be skill-specific in the sense that students with only comprehension practice will improve more in comprehension skills than students with only production practice and vice versa.

3b. Students with an equal amount of practice in both skills will perform at an intermediate level in both skills (cf., e.g., Anderson, 1993).

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

The Language

As will become clear later, it was important for this study to have rules that were completely independent of each other, that is, to avoid rules that cooccur, such as plural verb forms cooccurring with plural subject nouns, for instance. It was also necessary to have grammatical categories that correspond to characteristics of people and activities that can easily be presented visually. Furthermore, rules needed to apply without exception and present learners with binary options. Therefore, it was necessary to use an especially created miniature linguistic system (but with the characteristics of a natural language, unlike the artificial grammars generating letter strings used by cognitive psychologists). Such a miniature linguistic system offers the additional advantages that no students have any previous knowledge of the language or related languages and that no differences in integrative or instrumental motivation are to be expected. (See Hulstijn [this issue] for other studies of second language acquisition that have used (semi-)artificial languages.)

Autopractan, especially created for this study, is an agglutinative language with flexible word order. The noun receives number and case marking; the verb is marked for gender and instrumentality. The singular is marked with the suffix *-li,* the plural with *-to*. Nominative case is marked with the suffix *-fu,* accusative with *-ka*. Masculine and feminine are expressed by the suffixes *-ne* and *-pi,* respectively. The verb carries the suffix *-ra* when the action referred to is carried out by means of some instrument or tool, and the suffix *-so* when no tool is involved. Instrumental case is expressed by the suffix *-te*. Subject noun phrases can be dropped; Autopractan is a null-subject language. These rules comprise the entire morphosyntax of Autopractan.

The Autopractan lexicon contains 32 words, 16 nouns and 16 verbs. None of the lexemes resemble English words or words in the most commonly taught languages. See Appendix A for a list of the vocabulary items and their English equivalent; see Appendix B for some example sentences.

Subjects

The learners in this experiment were 61 paid volunteers, divided in three groups of about 20 each. Thirty-two participants were female and 29 were male. Most were undergraduates; a few were graduate students or staff members. The amount of payment depended on performance: Subjects were told at the beginning of the study that, at the end of the study and based on their performance throughout, they would receive $8 per hour if they scored in the top third of the group, $7 in the middle third, and $6 in the bottom third. As the rules to be taught involved gender marking and case marking, experience with Romance languages, German, Russian, Latin, or other commonly taught languages could have affected the results; therefore, participants in the three groups were matched on their experience with these languages.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

Design

The study had mostly a within-subjects design. Hypotheses 1 and 2 were tested by analyzing data of all the subjects over time (separately for comprehension and production, for reaction time and error rate, and for a single-task condition and a dual-task condition, the latter involving mental arithmetic during the language task). Hypothesis 3a was tested by comparing the same subjects on different (sets of) tasks. Hypothesis 3b, however, was tested by means of a between-subjects comparison.

Materials

A set of 400 color pictures were taken, corresponding to sentences in Autopractan; 16 were used for the initial vocabulary presentation, 256 for the practice and testing sessions, and 128 for the final testing session. All the pictures were digitized for use on a Macintosh computer.

A text about 10 computer screens long was written to present the Autopractan grammar rules listed in the section "The Language" to the subjects. See Appendix C for an example screen. The last part of the grammar presentation contained picture and sentence combinations.

Instruments

Three multiple-choice metalinguistic tests were developed to practice and test subjects' explicit rule knowledge; each test consisted of 10 items.

Each practice session consisted of both comprehension and production exercises and tests. For comprehension, sets of four pictures appeared on the screen with a sentence corresponding to one of the four pictures; the other three pictures would have required one or two different morphemes in the sentence. For production, one picture appeared with one or more blank fields to type in sentences or word endings. For both production and comprehension, sentences were scored as wrong as soon as there was any mistake. All sentences were either two or three words (six or nine morphemes) long.

The Modern Language Aptitude Test (MLAT) Form A (Carroll & Sapon, 1959) was administered to all subjects before the experiment in order to make sure that the different practice groups had the same level of language learning aptitude.

Procedure

The entire experiment, from the initial presentation of grammar and vocabulary to the final testing, took place on Power Macintosh 7200/75 computers; the software for all the components was written in HyperCard. Each individual participated in 22 sessions of an hour or less, spread over an 11-week period.

All students were taught the four grammar rules and 32 vocabulary items of Autopractan by means of a traditional presentation of grammar rules, along with a

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

presentation of the vocabulary accompanied with pictures to represent the people (nouns) and actions (verbs) involved. To make sure the grammar was thoroughly understood, a number of sentence and picture combinations were then presented with sets of questions about the inflectional morphemes and their relationship with the meaning represented by the pictures.

After going through this presentation twice (Sessions 1 and 2), the students were tested on their explicit knowledge of vocabulary and grammar (Sessions 3–6): Recall of all vocabulary items was tested in randomized order, and a fill-in-the-blanks test about explicit grammar rule knowledge was administered. In Sessions 3–5, a different 10-item test was used in each session. The participants received explicit feedback on every response, explaining in detail for each error why the response was wrong. In Session 6, a combined test of all 30 previously used items was administered. At that time, all students demonstrated complete knowledge of both vocabulary and grammar, in the sense that they made no mistakes (in most cases) or just one or two (most likely due to distraction, as was often clear when subjects said "oops" immediately after choosing an answer).

Afterward, the subjects were assigned to one of three practice conditions. Group A practiced the rules for number of the noun and instrumentality of the verb in comprehension and the rules for case marking on the noun and gender marking on the verb in production. For Group B it was the other way around: Participants in Group B practiced the rules for number of the noun and instrumentality of the verb in production and the rules for case marking on the noun and gender marking on the verb in comprehension. Group C practiced all four rules in both comprehension and production, but with half as many items each as Groups A and B, so that the total number of items was the same for all three groups. The subjects practiced comprehension by selecting one of a set of four pictures that matched a sentence displayed in the middle of the screen; the four pictures were always chosen so that the two grammar features being practiced were critical to choosing the correct picture (number marking on the noun and instrumentality marking on the verb for all the sets of pictures shown to Group A; case marking on the noun and gender marking on the verb for all the sets of four pictures shown to Group B; and one of these two combinations for each set of four pictures shown to Group C, depending on the set). Production practice took the form of typing in the sentence corresponding to a picture displayed on the screen. Again, two grammar rules were critical to giving a correct answer (case marking on the noun and gender marking on the verb for all the pictures shown to Group A; number marking on the noun and instrumentality marking on the verb for all the pictures shown to Group B; one of these two combinations for each picture shown to Group C, depending on the picture). Having a null-subject language made it possible never to present sentences with subject–object contrasts in comprehension for Group A and in production for Group B (by presenting sentences without subject for comprehension and prompting for sentences without subject in production). The other contrasts were simply avoided for the group in question by making sure all pictures or picture sets represented only masculine subjects, or only singular nouns, or only actions carried out without a tool. Having complete control over what grammatical contrasts appeared in a given picture was

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

necessary to make sure that the participants would only practice two out of four grammatical contrasts in production and the other two in comprehension.

In each of the 15 practice sessions (Sessions 7–21), subjects worked through eight blocks of items. Four blocks for practice were always followed by four blocks for testing. Within each group of four, there were two blocks of production items and two blocks of comprehension items; for both comprehension and production, one of the two blocks was single-task and one was dual-task. In the dual-task condition, subjects saw a number between 100 and 1,000 in the middle of the screen before the picture(s) appeared. While they were carrying out the comprehension or production task, they heard beeps at irregular intervals.[2] The secondary task consisted of memorizing the three-digit number, counting the beeps, and subtracting the number of beeps from the original number immediately after giving a response to the comprehension or production item.[3] The outcome of the subject's computation was recorded along with the answer to the main task and the reaction time.[4] The reason for the dual-task condition was to allow for an assessment of how automatized the language skills were: The more a skill is automatized, the less it should interfere with a secondary task or show interference from it; that is, the more it is automatized, the less difference is expected between the single-task and dual-task conditions.

The four testing blocks of picture and sentence combinations each consisted of 16 items in all 15 practice sessions; the four practice blocks always consisted of 8 items each. The order in which pictures or sets of pictures were presented was randomized for each block; so was the choice of the combination of morphemes out of the possible set of four combinations, as well as the position on the screen for the one correct and three incorrect pictures in the comprehension blocks. In the first session, the practice items were in English instead of Autopractan. In this way, the subjects had the opportunity to get used to the testing formats before ever having practiced a single Autopractan sentence in comprehension or production. This set-up allowed for the collection, in the second half of the first practice session, of ideal baseline data for Autopractan comprehension and production skills: The learners had not had any previous practice of the comprehension and production skills themselves but had practiced the ancillary skills (e.g., dealing with the keyboard and screen layout) and the relevant declarative knowledge (Autopractan vocabulary and grammar as taught in Sessions 1 and 2 and practiced in declarative form in Sessions 3–6). During the practice blocks, feedback was provided; each error triggered an automatic explicit explanation of what grammatical categories in the answer were wrong and why, for both the production and comprehension tasks.[5] No feedback was given during the testing blocks.

For the production tasks, there was a slight difference in format between the practice and the testing blocks. In the four practice blocks in each session, subjects had to type in the complete sentence corresponding to the picture on the screen. In the testing blocks, the lexemes were provided and the subjects only needed to type in the grammatical morphemes. In this way, the practice was maximally representative of language production, but the reliability and validity of the test items was improved by avoiding all problems of vocabulary knowledge and minimizing the likelihood of typographical errors.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

The four practice blocks always came before the four testing blocks, but the order of comprehension versus production blocks and single-task versus dual-task blocks was counterbalanced over sessions (even vs. uneven session numbers) and subjects (even vs. uneven subject codes). Thus, when subjects with an uneven code started with a certain order of blocks in Session 1, then subjects with an even code would have the opposite order for both comprehension versus production and single-task versus dual-task in that session, and the subjects with an uneven code number would have the opposite order in Session 2, and so on.

To summarize, all subjects were taught the same rules explicitly and then received the same amount of practice and exposure for each rule. All subjects also had the same amount of practice in both skills and therefore also the same amount of experience with both practice and testing formats. The only difference between the three groups was the amount of practice of specific rules in specific skills and formats. Thus, the design allows for testing the specificity of procedural skill (comprehension vs. production) acquired as a result of systematic practice of that skill after all subjects had achieved the same quantity and quality of explicit knowledge.

After the 15 sessions of practicing and testing specific combinations of rules and skills, a final session (Session 22) consisted of two parts. In one half of the final session, the subjects were tested, with a new set of sentences and pictures, on the usual combinations of rules and skills (without warm-up this time) ("same" condition). In the other half, they were tested on the opposite combination: The rules that were previously practiced and tested in comprehension were now tested in production, and vice versa ("reverse" condition). Whether subjects were first administered the items in the "same" and then the "reverse" condition or the other way around was counterbalanced over subjects (low vs. high numbers in subject codes). The order of comprehension versus production blocks and single-task versus dual-task blocks was also counterbalanced over subjects (even vs. uneven subject codes). It should be stressed here that the subjects did not engage in any new tasks at this stage and were not tested on any new rules, only on task and rule combinations they had not encountered before.

RESULTS

The Learning Curve

Figures 1–4 show the mean performance of all participants as a function of practice for both comprehension and production in terms of reaction times and error rates.

Reaction times for comprehension are shown in Figure 1. Performance under single-task and dual-task conditions is very similar, except for the first session, in which it is slightly better for the single-task than for the dual-task condition.

Reaction times for production are presented in Figure 2. Reaction times in both single-task and dual-task conditions show a very slow decline. Performance for the dual-task conditions is virtually indistinguishable from that of the single-task condition.

Figure 3 shows the number of errors for comprehension. The curves are somewhat

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

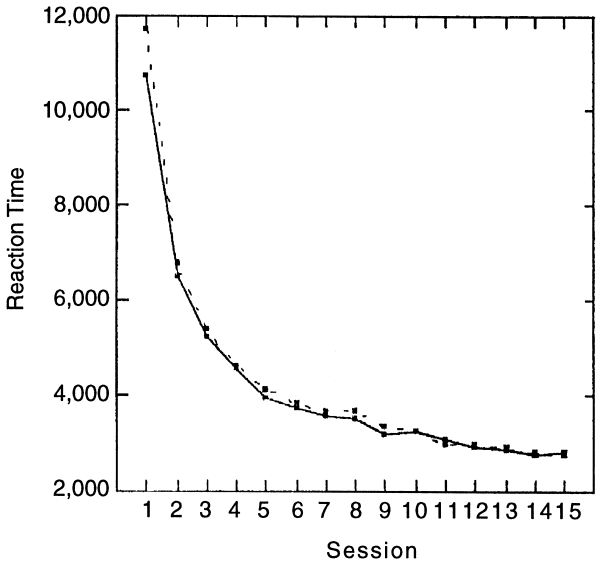


Figure 1. Reaction time (ms) in the comprehension task as a function of practice (solid line = single-task condition; dashed line = dual-task condition).

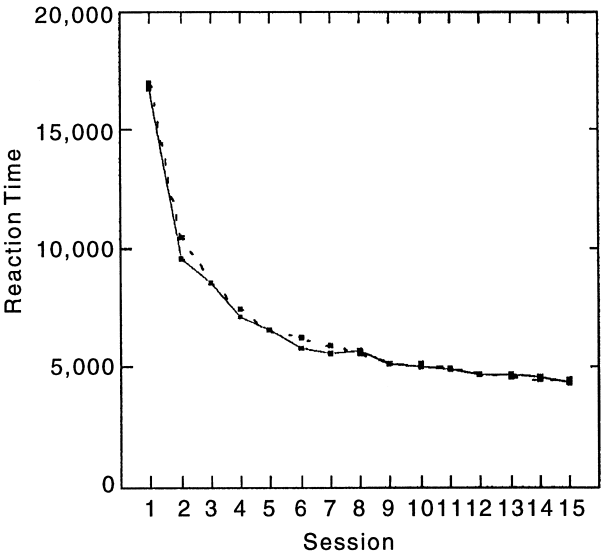


Figure 2. Reaction time (ms) in the production task as a function of practice (solid line = single-task condition; dashed line = dual-task condition).

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

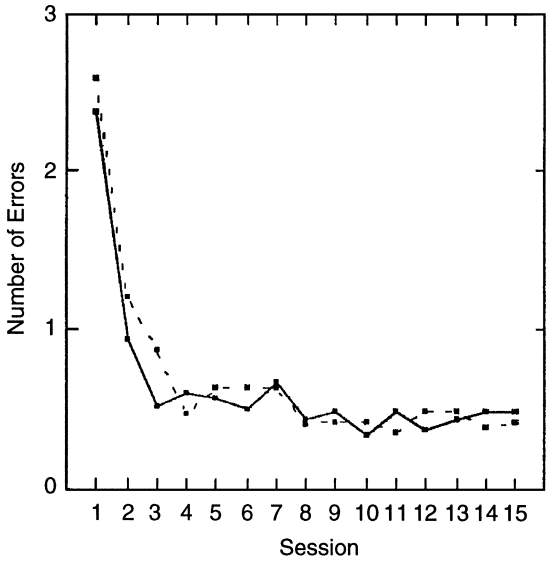


Figure 3. Number of errors in the comprehension task as a function of practice (solid line = single-task condition; dashed line = dual-task condition).

bumpy for both conditions. Subjects in the single-task condition did better than those in the dual-task condition for the first three sessions. After that, there appears to be no pattern to the relative performance of single-task and dual-task conditions.

Figure 4 shows the number of errors for production. The single-task and dual-task conditions yielded largely similar results, except for the first session. Here subjects did better in the single-task condition, too.

Transfer Across Skills

Table 1 presents the means and standard deviations for the various measures of performance in the "same" condition of the final test (in which the participants used the rules in comprehension that they had always practiced in comprehension and used the rules in production that they had always practiced in production) and in the "reverse" condition (in which they had to use each rule in the opposite skill compared to what they were used to), along with significance statistics for the difference between the two conditions. For both production tasks and for the single-task condition of the comprehension task, the difference between same and reverse, as measured by a paired-samples t test, is always significant at $p < .05$, and usually at $p < .01$ level, for both reaction times and error data. For the comprehension and dual-task combination, however, only reaction time shows a significant difference for the two conditions.

Table 1 also presents data from the mixed-practice group (C) for comparison,

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

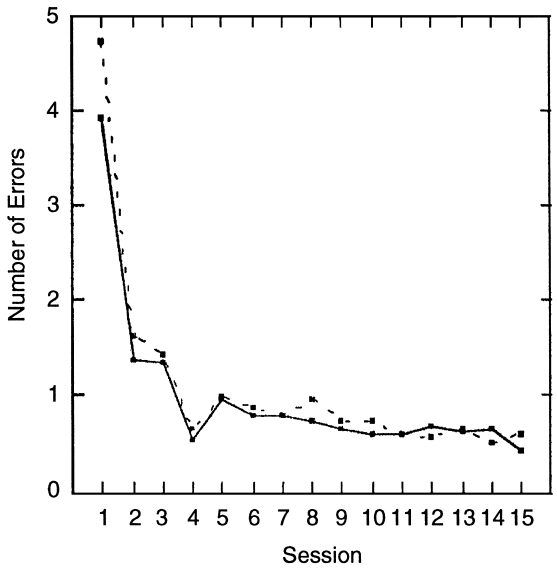


Figure 4. Number of errors in the production task as a function of practice (solid line = single-task condition; dashed line = dual-task condition).

along with significance data for the comparisons between the performance of this group on the final test and that of the specialized-practice groups (A and B) under the "same" and "reverse" conditions. After determining that no significant difference existed between Groups A and B on the one hand and Group C on the other in terms of MLAT scores, and therefore that an analysis of covariance was not necessary, a set of independent-sample *t* tests was carried out for the various performance measures. As the last four columns in Table 1 show, the differences between Group C and the same condition are never significant, except for the dual-task condition of the comprehension task. The differences between Group C and the reverse condition are always significant, with the exception again of the comprehension and dual-task combination.

DISCUSSION

Hypothesis 1

The hypothesis of gradual automatization as a function of practice was largely confirmed. For both reaction time and error rates, for both production and comprehension, and for both single- and dual-task conditions, a clear and gradual drop-off takes place, which slows down over time, from a dramatic decrease between Sessions 1 and 2 to an almost imperceptible decline for the last few sessions. These results confirm the findings of a pilot study with the single-task condition only, with fewer

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

Table 1. Performance measures for the "same" and "reverse" conditions of the final test

	Same (Groups A/B) ($n = 41$)		Reverse (Groups A/B) ($n = 41$)		Group C (Same/ Reverse) ($n = 20$)		Difference Same vs. Reverse		Difference C vs. Same		Difference C vs. Reverse	
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*	*p*	*t*	*p*	*t*	*p*
Comprehension												
Single task												
Reaction time (ms)	3,641	1,923	5,767	4,039	3,406	1,199	4.01	.000	0.58	.562	3.45	.001
Errors (out of 16)	0.63	0.77	1.42	1.25	0.75	0.72	3.11	.003	–0.58	.566	2.64	.011
Dual task												
Reaction time (ms)	3,917	1,939	4,800	1,837	3,301	962	2.27	.029	1.66	.103	4.18	.000
Errors (out of 16)	1.24	1.58	1.07	1.17	0.65	0.61	–0.52	.605	2.11	.039	1.86	.068
Production												
Single task												
Reaction time (ms)	4,964	2,126	8,325	4,014	5,182	1,319	3.48	.000	–0.39	.697	3.15	.000
Errors (out of 16)	1.00	1.84	2.68	3.75	0.60	0.50	2.09	.043	1.29	.202	3.50	.001
Dual task												
Reaction time (ms)	5,506	2,531	7,591	2,845	4,997	876	3.09	.004	1.15	.255	4.68	.000
Errors (out of 16)	0.93	2.03	2.93	3.76	0.63	0.63	2.69	.010	.87	.388	3.82	.000

subjects and fewer sessions (DeKeyser, 1996). The curves are very smooth for reaction time and bumpier for the error data. It is well known that error data are less reliable (especially with short tests) and therefore harder to fit to a curve than reaction time data (cf., e.g., Anderson, 1995). The difference between the single-task and dual-task conditions did not decline as slowly and gradually as expected. The difference is small in most cases, except for the error rate in the first two or three sessions. A two-way repeated-measures analysis of variance (first vs. last session × single-task vs. dual-task condition) yielded no significant interaction effect, only a main effect for session. It appears that the arithmetic task was too easy to interfere much with the language task once subjects had a modest amount of practice. As a result, the interaction is in the expected direction, but it is not significant.

Hypothesis 2

The second hypothesis is more specific. Longitudinal studies of skill acquisition in various domains, including perceptual-motor as well as cognitive skills, have yielded data that follow a power law (Newell & Rosenbloom, 1981). That means that when practice (expressed in terms of time elapsed or number of learning sessions) is x and reaction time or error rate is y, then $y = x^n$. When empirical data that fit a power function are plotted on a logarithmic scale ($\log(y)$ vs. $\log(x)$), they come close to yielding a straight line; the correlation between $\log(y)$ and $\log(x)$ is close to 1. Other functions exist that are characterized by a gradual decline, such as an exponential function ($y = e^x$). If a function is exponential, then a plot of $\log(y)$ versus x is a straight line and the correlation of $\log(y)$ with x is 1, but Newell and Rosenbloom (1981) show that power functions give a consistently better fit than exponential functions for data that reflect the effect of practice, in the sense that the correlation between $\log(y)$ and $\log(x)$ is very large (typically above 0.95) and higher than the correlation between $\log(y)$ and x.

Our data for reaction time show a very close fit to a power function: The correlations between log(reaction time) and log(session number) vary from –.983 for production in the single-task condition to –.991 for production in the dual-task condition (see Table 2). The corresponding coefficients for an exponential function ($r_{\log(y)x}$) are –.872 and –.891. Visual inspection of the corresponding graphs showed that, even with correlation coefficients as high as –.991, there was still a slight curvature in the spread of the data points around the regression line, mainly due to the fact that the data for the first session have an unduly strong influence on the position of the regression line. Therefore, following Newell and Rosenbloom (1981), the data were analyzed again with the first session removed. Although this had only a slight influence on the correlation coefficients (now ranging from –0.990 to –.996), the data points are now symmetrically distributed around the regression line. In Figures 5–8 the regression line drawn is the closest fit for the data points from Sessions 2–15. The data point for Session 1 has been plotted, too, however, in order to show how it deviates from the pattern for the subsequent sessions. These graphs show even more convincingly than the correlation coefficients by themselves that the data are a perfect fit for a power function in the sense that the data points are virtually on

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

Table 2. Correlations between number of practice sessions and performance measures as measured by $r_{\log(y)\log(x)}$ and $r_{\log(y)x}$

	All Sessions		Sessions 2–15	
	$r_{\log(y)\log(x)}$	$r_{\log(y)x}$	$r_{\log(y)\log(x)}$	$r_{\log(y)x}$
Reaction time				
Comprehension				
Single task	–.987	–.878	–.992	–.932
Dual task	–.987	–.881	–.994	–.941
Production				
Single task	–.983	–.872	–.990	–.934
Dual task	–.991	–.891	–.996	–.942
Error rate				
Comprehension				
Single task	.783	.570	.744	.636
Dual task	.849	.647	.847	.726
Production				
Single task	.807	.604	.825	.741
Dual task	.809	.607	.872	.788

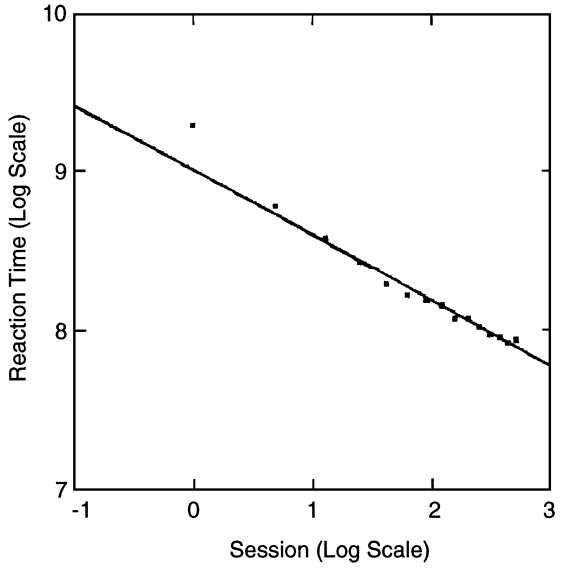


Figure 5. Reaction time as a function of number of practice sessions (Session 1 removed from analysis but not from graph): comprehension task/single-task condition.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

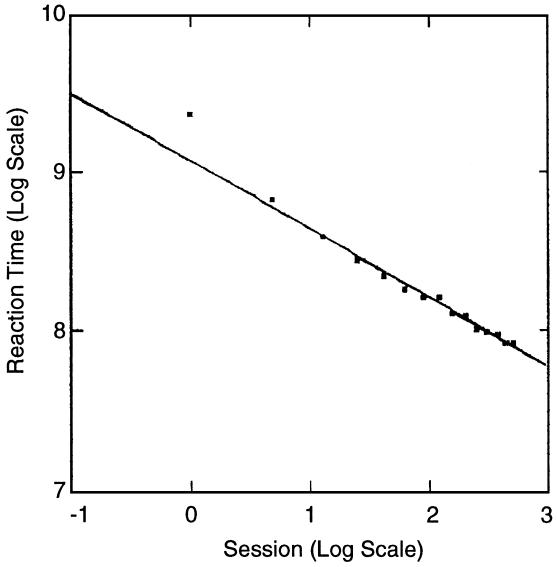


Figure 6. Reaction time as a function of number of practice sessions (Session 1 removed from analysis but not from graph): comprehension task/dual-task condition.

the regression line and do not show a consistent residual pattern. The graphs for the corresponding exponential function (not included in this article) consistently show a nonsymmetrical distribution of the data points, above the regression line for the first and last few sessions, below in the middle. In other words, the exponential function is a less good fit, not just because of scatter around the regression line, but because it fails to capture the exact shape of the learning curve as well as the power function does.

It is important to note that the data from Session 1 are not just being excluded in order to provide a better fit for the theoretical functions. In fact, it follows from the model of skill acquisition outlined in the literature review that initial practice of a task has very different effects (proceduralization of declarative knowledge) compared to subsequent practice (automatization). Anderson and Fincham (1994), in particular, argue for the qualitatively different nature of the initial practice and the speed with which proceduralization takes place.[6]

The data for error rates show a less good fit, as is to be expected from a cursory visual inspection of the graphs in Figures 1–4, but even there the power function provides a consistently better fit than the exponential function (see Table 2).[7]

In summary, our data show that practice in the use of second language grammar rules, whether through comprehension or production tasks, has exactly the same effect on learning as it does in the acquisition of other cognitive skills. This strongly suggests that the same mechanisms are involved and that the ability to comprehend

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

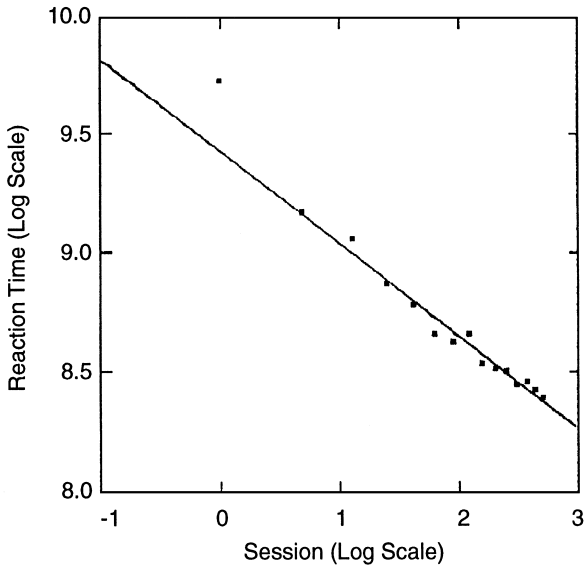


Figure 7. Reaction time as a function of number of practice sessions (Session 1 removed from analysis but not from graph): production task/single-task condition.

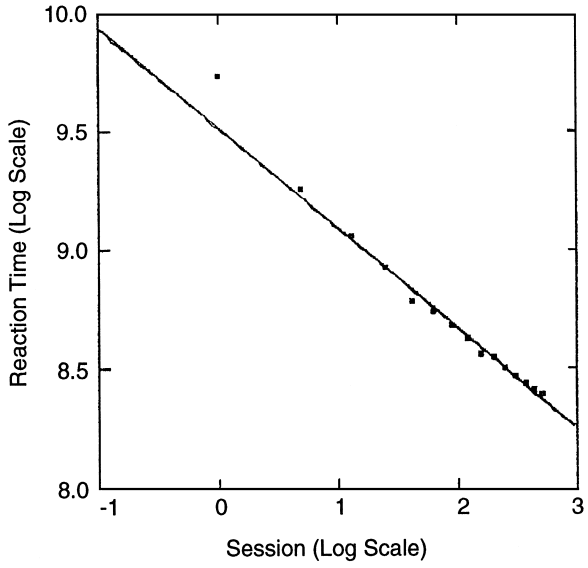


Figure 8. Reaction time as a function of number of practice sessions (Session 1 removed from analysis but not from graph): production task/dual-task condition.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

or produce sentences in a second language is not necessarily acquired through the implicit mechanisms of a separate mental module (as is generally accepted for first language acquisition).

Hypothesis 3

If the hypothesis that the effect of practice is skill-specific is correct, then students with comprehension practice only should improve more in comprehension skills than students with production practice only and vice versa, whereas students with an equal amount of practice in both skills should perform at an intermediate level in both skills.

This hypothesis was largely confirmed, too, by the data from the final test (with all new pictures). The "same" and "reverse" condition columns in Table 1 show that the subjects almost always performed worse in the "reverse" condition than in the "same" condition in terms of both reaction times and error rates, as the participants in the pilot study (DeKeyser, 1996) did. For seven out of eight combinations of task, task condition, and criterion measure, the difference is rather large (especially for production, in which reaction times and error rates are roughly 1.5–3 times as high in the "reverse" condition compared to the "same" condition), and it is significant at $p < .05$ or better in a pairwise t test. For error rate in the comprehension task under dual-task conditions, the difference is not significant, and it is actually in the opposite direction from what was expected. This one failure to reach significance is probably due to the low reliability of the comprehension task (16 items, four-way multiple-choice).

It was further predicted that the results for Group C (mixed-practice group) would always be in-between the results for the "same" and "reverse" conditions for the specialized-practice groups (A and B). Table 1 shows that this was not the case: Six out of the eight measures even show a slight (nonsignificant) advantage for the mixed-practice group over the "same" condition of the specialized-practice group. As predicted, however, the mixed-practice group always shows a distinct advantage over the reverse condition for the single-practice group. This difference is significant at $p < .01$ for all measures except the two comprehension error rate measures (single-task and dual-task conditions), in which it narrowly fails to reach significance (at $p < .01$ and $p < .05$, respectively). Here, again, the failure to reach significance for this one particular pair of measures is probably due to the low reliability of the multiple-choice task.

Overall, our data lend strong support to the hypothesis that practice has a skill-specific effect in the sense that performance in comprehension or production is severely reduced if only the opposite skill was practiced. This is contrary to the idea of linguistic competence acquired through comprehension being equally available for production and comprehension. At the same time, however, the data for the mixed-practice group show that, once a certain level of practice in a skill has been reached, additional practice in that same skill (in the sense of more items over the same time period) does not yield better results. It appears that, within a given session, the learning effect tends to asymptote after a certain numbers of items. This may be

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

another illustration of the same power law that was shown in the change from session to session.[8]

It should be stressed that these findings go beyond those of VanPatten and associates (Cadierno, 1995; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz 1995) on the one hand and DeKeyser and Sokalski (1996) on the other. Whereas the disagreement between those studies concerns the short-term effect of practice before any automatization has taken place, the present findings address the long-term effect of systematic practice, once the automatization process has reached asymptote or come close, that is, once learning has bottomed out in terms of both reaction time and error rate.

CONCLUSIONS AND IMPLICATIONS

This study has shown that the learning of second language grammar rules can proceed very much in the same way that learning in other cognitive domains, from geometry to computer programming, has been shown to take place. The learning curves observed in this study follow the same power law as for those other skills. As has been found in several studies in other domains, the first practice session presented an exception to the pattern followed by the other sessions. This evidence supports the model of skill acquisition that posits that during initial practice declarative knowledge is turned into qualitatively different procedural knowledge and that subsequently a much slower process of gradual automatization takes place, which requires little or no change in task components, only a quantitative change within the same components.

Evidence has also been presented to the effect that practice leads to highly specific skills. Just as was found previously in studies of computer programming, this experiment has shown that large amounts of practice in comprehension and production of second language grammar rules lead to spectacular improvements in the skill practiced but have far less impact on the reverse skill. When subjects had to use rules in comprehension that they had previously practiced in production only, and vice versa, the reaction time and error rate often went up by a factor between 1.5 and 3, especially for production.

The way the study was designed makes it clear that the two main findings presented here, the gradual improvement following the power law and the skill specificity of the learning process, are language learning phenomena and not just task habituation effects. Not only did subjects get practice in the mechanics of the various task formats before the first test data were collected but, more importantly, the students' performance between the "same" and "reverse" conditions of the final test shows a big difference between rules practiced in the skill tested and rules only practiced in the other skill. As there was absolutely no difference between these two conditions for the group as a whole in terms of task format, pictures, or sentences, the differences in performance are clearly due to the degree to which grammar rules had been practiced.

In terms of teaching methodology, findings such as these suggest that there is a role for systematic practice of specific rules for specific skills in the second language

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

curriculum. This does not imply that practice should be as focused and repetitive as it was in this study, not even that a structural syllabus should be used. It does imply, however, that the sequence of explicit rule learning, followed by a short period of activities focused on using explicit knowledge during performance of the target skills, and finally by a long period of repeated opportunities to use that knowledge, is likely to yield knowledge that is highly automatized.

Three caveats are in order, though. First of all, proceduralization and automatization may not be equally successful for all learners and all rules. Second, no claim is being made here that the automatized knowledge documented in this study is equivalent to the implicit knowledge typically acquired in the native language (as an artificial language was used, no native speaker comparisons could be made). It remains to be seen whether more prolonged practice or practice that encourages less focus on form could induce such fundamental change. Third, no specific recommendations for teaching methodology should be made before the findings of this study have been put to the test in the more ecologically valid context of real second language classrooms.

In terms of research methodology, this study has shown that research in a laboratory context allows forms of intensive, fine-grained, and highly controlled data collection that, in turn, allow for more precise testing of more specific hypotheses than would be possible in a more naturalistic context. Findings from such laboratory research should not be overgeneralized, just as findings from biochemical studies should not be used to recommend large-scale use of new drugs without careful clinical studies. By the same token, however, it can be ill-advised to test complex hypotheses in highly variable and poorly understood classroom contexts, just as it would be counterindicated to test out new drugs on patients before their biochemical properties are at least known in principle. Encouraged by the results of the present study, which shows that second language grammar rules can be learned and automatized like rules in other domains, at least under laboratory conditions, we are now planning a number of classroom experiments on the long-term effect of practice activities designed according to Anderson's model of skill acquisition.

Finally, the use of an artificial language has been shown to be useful for various reasons. It not only has the obvious advantages of controlling previous knowledge, outside exposure, or integrative motivation, but perhaps, more importantly, it allows the researcher to remove a number of linguistic constraints on an already complex experimental design, which otherwise may become completely unmanageable.

NOTES

1. I am using ACT here as a general term for the original ACT model (Anderson, 1976), the ACT* (pronounced ACT–star) version introduced in Anderson (1983), and the ACT-R (ACT–Rational) version introduced in Anderson (1993).

2. During practice, for comprehension, the first beep occurred after 3 seconds; subsequent beeps followed after intervals varying between 5 and 6 seconds in length. For production, the first beep occurred after 3 seconds, and subsequent beeps after intervals of 5.5–6 seconds. During testing, for comprehension, the first beep occurred after 3.5 seconds; the intervals for subsequent beeps varied between 5 and 6 seconds in length. For production, the first beep occurred after 3 seconds; subsequent beeps occurred after 2–3-second intervals. These intervals were determined on a trial-and-error basis during the piloting.

3. Arithmetic tasks are commonly used as secondary tasks in experimental psychology to limit the amount

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

of cognitive resources that participants can allocate to the primary task when both are executed simultaneously (see, e.g., Pashler, 1994). Counting backward in steps of seven from a three-digit number, for instance, is a commonly used procedure. During the initial stages of piloting the Autopractan experiment, I found that counting in steps of more than one was unduly difficult in this case and that simply asking subjects to subtract each time they heard a beep was likely to lead to a situation where some subjects did this literally, and others would count the number of beeps, and then subtract that number from the original number. The procedure adopted here was meant to avoid these problems.

4. Reaction time for the comprehension tasks was measured as the time elapsed between presentation of the stimulus sentences and pictures and hitting the button corresponding to one of the four picture choices. For the production tasks, reaction time was the time elapsed between presentation of the picture stimulus and hitting the return button after finishing typing in the sentence. Reaction time for the arithmetic task was not measured, nor was the correctness of the arithmetic task taken into account for the data analysis. Subjects were motivated to do well on this task, however, because the correctness of the arithmetic task was recorded and taken into account to determine their payment.

5. In the comprehension format, errors were detected by a comparison of a series of grammar codes for the sentence in question with the grammar codes for the picture chosen. In case of mismatch for one of the codes, participants were told that there was a problem with the corresponding category (i.e., gender, number, instrumentality, or case marking). Otherwise, they were told the picture was correct. In the production format, the program scanned the input for all the required words in any order. When an error was detected, the student was told what the expected word was, along with a description of its grammatical features. In the metalinguistic format (Sessions 3–6) ad hoc feedback was given to each specific choice made in the multiple-choice format for the grammar items. In case of errors in the vocabulary test, the right word was presented to the student. In this vocabulary test, the ERRATA program developed by Robert Hart (University of Illinois at Urbana–Champaign) was used for comparing input words to expected answers.

6. To determine how fast proceduralization takes place, a separate analysis was performed on the data from the subjects for whom the dual-task condition preceded the single-task condition in the first session (i.e., half of the subjects, as the order of conditions was counterbalanced). For these subjects, the results for the dual-task condition show an even more dramatic deviation for the first session (the first data point is almost twice as far removed from the regression line as for the group as a whole), whereas for the single-task condition the first data point falls exactly on the regression line. This shows that the proceduralization process is complete not only by the end of the first session but already by the end of the first 16-item block for each skill.

7. The coefficients in the bottom half of Table 2 refer to the correlation with number correct out of 16 instead of number of errors out of 16, contrary to the previous tables and graphs, which present number of errors. This was done to avoid having to take the logarithm of 0 in each case where there were no errors, which would have led to many missing values.

8. I thank Nick Ellis for pointing this out to me.

REFERENCES

Alanen, R. A. (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 Hawaii Press.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.

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

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94,* 192–210.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105,* 165–180.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1995). *Learning and memory. An integrated approach*. New York: Wiley.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 1322–1340.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal, 79,* 179–193.

Carroll, J. B., & Sapon, S. (1959). *Modern Language Aptitude Test. Form A*. New York: The Psychological Corporation.

Chastain, K. (1971). *The development of modern language skills: Theory to practice*. Philadelphia: Center for Curriculum Development.

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

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379–410.
DeKeyser, R. (1996). Exploring automatization processes. *TESOL Quarterly, 30,* 349–357.
DeKeyser, R. (in press). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom language acquisition*. New York: Cambridge University Press.
DeKeyser, R., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning, 46*, 613–641.
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.
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, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27,* 91–113.
Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory and Cognition, 11,* 565–574.
Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22,* 473–492.
Griffiths, R. (1991). Pausological research in an L2 context: A rationale and review of selected studies. *Applied Linguistics, 12,* 345–364.
Hart, R. (1994). *ERRATA: Response analysis and error diagnosis tools* (Tech. Rep. No. LLL-T-23-94). Language Learning Laboratory, University of Illinois at Urbana–Champaign.
House, J. (1996). Developing pragmatic fluency in English as a foreign language. *Studies in Second Language Acquisition, 18,* 225–252.
Hulstijn, J. H. (1997). Introduction. Second language acquisition research in the research lab: Possibilities and limitations. *Studies in Second Language Acquisition, 19,* 131–144.
Hulstijn, J., & 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.
Johnson, K. (1996). *Language teaching and skill learning.* Oxford: Blackwell.
Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Englewood Cliffs, NJ: Prentice-Hall.
Krashen, S. D. (1985). *The input hypothesis*. New York: Longman.
Krashen, S. D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–77). London: Academic Press.
Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.
Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 883–914.
Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 1022–1050.
Logan, G. D., Taylor, S., & Etherton, J. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 620–638.
Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229–252). New York: Cambridge University Press.
McLaughlin, B. (1987). *Theories of second-language learning.* London: Edward Arnold.
McLeod, B., & McLaughlin, B. (1986). Restructuring or automaticity? Reading in a second language. *Language Learning, 36,* 109–123.
Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 28–49). Tübingen, Germany: Gunter Narr.
Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.
O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.
Paulston, C. B., & Bruder, M. N. (1976). *Teaching English as a second language: Techniques and procedures*. Cambridge, MA: Winthrop.
Raupach, M. (1987). *Procedural learning in advanced learners of a foreign language*. (Linguistic Agency University of Duisburg Paper B167). Duisburg, Germany: Universität Duisburg Gesamthochschule.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

Rivers, W. M. (1964). *The psychologist and the foreign language teacher*. Chicago: University of Chicago Press.
Robinson, P. (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., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition, 15,* 413–438.
Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357–385.
Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. *Human Factors, 30,* 539–566.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Part I. Detection, search, and attention. *Psychological Review, 84,* 1–66.
Scott, V. M. (1989). An empirical study of explicit and implicit teaching strategies in French. *The Modern Language Journal, 73,* 14–22.
Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14,* 369–385.
Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–196.
Spada, N., & Lightbown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15,* 205–224.
Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84–119.
VanPatten, B., & Cadierno, T. (1993a). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15,* 225–243.
VanPatten, B., & Cadierno, T. (1993b). Input processing and second language acquisition: A role for instruction. *The Modern Language Journal, 77,* 45–57.
VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. Eckman et al. (Eds.), *Second language acquisition theory and pedagogy* (pp. 169–185). Mahwah, NJ: Erlbaum.
White, L., Spada, N., Lightbown, P., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12,* 416–432.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

APPENDIX A

AUTOPRACTAN–ENGLISH DICTIONARY

balit, V	"hit," "punch," "beat"	*kupif,* N	"sports fan"
barat, N	"soldier"	*ladet,* N	"waiter"
barip, V	"tickle"	*limot,* V	"push"
bimet, N	"scout"	*mator,* V	"watch," "look at"
biruf, N	"sailor"	*nabok,* N	"worker"
bomap, N	"rugbyplayer"	*nemok,* V	"strangle"
bulot, V	"point at"	*patek,* V	"pull," "drag"
butak, N	"motorcyclist"	*pemek,* N	"clown"
fatis, V	"feed"	*potam,* N	"doctor," "physician"
fatuk, V	"aim at," "shoot at"	*pumaf,* V	"spank"
fitop, N	"king," "queen," "monarch"	*samek,* V	"examine"
fokap, N	"student"	*selek,* N	"cowboy"
fumas, V	"block," "stop," "halt"	*sipuk,* V	"measure"
kabom, N	"cook"	*sotot,* N	"viking"
kapit, V	"scare"	*titom,* N	"backpacker"
kotif, V	"poke"	*topim,* V	"trip"

Note: N = noun; V = verb.

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

APPENDIX B

SOME SAMPLE SENTENCES IN AUTOPRACTAN

(1) *selek-li-fu* *fatuk-ne-ra* *fitop-li-ka*
cowboy-SG-SUBJ shoot-MASC+INST king-SG-OBJ

or

fitop-li-ka *fatuk-ne-ra* *selek-li-fu*
king-SG-OBJ shoot-MASC+INST cowboy-SG-SUBJ

"The cowboy is shooting at the king." (with a gun, for instance)

(2) *potam-li-fu* *samek-pi-so* *titom-li-ka*
doctor-SG-SUBJ examine-FEM-noINST backpacker-SG-OBJ

or

titom-li-ka *samek-pi-so* *potam-li-fu*
backpacker-SG-OBJ examine-FEM-noINST doctor-SG-SUBJ

"The doctor [fem.] is examining the backpacker." (without an instrument)

(3) *barat-to-fu* *mator-ne-so* *kabom-to-ka*
soldier-PL-SUBJ look-MASC-noINST cook-PL-OBJ

or

kabom-to-ka *mator-ne-so* *barat-to-fu*
cook-PL-OBJ look-MASC-noINST soldier-PL-SUBJ

"The soldiers are looking at the cooks." (without an instrument)

(4) *mator-pi-ra* *bimet-to-ka*
look-FEM-INST scout-PL-OBJ

or

bimet-to-ka *mator-pi-ra*
scout-PL-OBJ look-FEM-INST

"They [fem.] are looking at the girlscouts." (with binoculars, for instance)

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press

APPENDIX C

SAMPLE SCREEN FROM THE GRAMMAR PRESENTATION

Nouns in Autopractan differ depending on whether they are subject or object of the sentence—in other words, depending on whether they express who does the action or who undergoes it. The ending *-fu* on a noun marks a grammatical subject, the ending *-ka* marks a grammatical object. Compare the following:

pemek-li-fu	*mator-ne-so*	*fokap-li-ka*
the (male) clown (SUBJ)	is watching	the student (OBJ)
fokap-li-fu	*mator-ne-so*	*pemek-li-ka*
the (male) student (SUBJ)	is watching	the clown (OBJ)

As can be seen, a noun normally occurs with two endings in Autopractan, one for singular (*li*) versus plural (*to*) and one for subject (*fu*) versus object (*ka*).

https://doi.org/10.1017/S0272263197002040 Published online by Cambridge University Press