The Effect of a Self-Selected Extensive Reading Program on Nara University Students' English Writing Proficiency: A Flawed Study

James SWAN

College of Liberal Arts

ABSTRACT

Seven sections of Nara University Freshman and Sophomore students were given extensive reading classes in place of half the usual intensive reading curriculum. Narrative writing samples were taken at the beginning and the end of the academic year and scored for fluency and accuracy with an open-ended yet highly sensitive objective rating technique (Ishikawa, in press). For the aggregate group, the mean scores of paired t-tests on all criteria rose over time. Four of the seven sections were also sampled at the mid-year point; comparisons between the sampled and unsampled sections suggest the existence of a beneficial practice effect. The data presented here are contaminated by operational errors in the conduct of the experiment, but the results are nevertheless construed as support for Krashen's "Monitor Model" in general and the "Input Hypothesis" in particular, and as empirically validating the usefulness of Ishikawa's technique.

BACKGROUND

In recent years, probably no educational theoretician has inspired more active research and debate among language teachers around the world than Prof. Stephen D. Krashen. After several productive years conducting and publishing basic psycholinguistic research, he burst into prominence in the language teaching field with his theory of language acquisition, which he called the "Monitor Model of Adult Language Performance." Krashen's intuitively attractive description of the language acquisition process thrust him into the language teaching limelight, stimulating much controversy and confirmatory research.

Received September 12, 1994

Krashen's model consists of five interrelated hypotheses. Because a basic understanding of Krashen's theory is essential to an understanding of my project, I present these five hypotheses here very briefly, at the considerable risk of vastly oversimplifying them. They are:[1]

(1) *The Acquistion-Learning Hypothesis* , which asserts basic, crucial differences between these two processes. According to this hypothesis, human beings subconsciously acquire language simply by understanding messages, whereas we learn language by formal study of its elements and described rules. Consciously learned language does not necessarily become subconsciously acquired language;

(2) *The Natural Order Hypothesis*, which claims that language forms are acquired in a natural developmental sequence, regardless of the sequence formally studied. Acquisition is biologically determined by the brain's readiness for the next natural stage, rather than by artificially determined levels or sequences of formal instruction;

(3) *The Monitor Hypothesis*, probably the most controversial of the five hypotheses, which asserts that language production is a function of acquired language, not learned language. Learned language is not in itself sufficient for initiating production, but only serves as a "mental editor" or "monitor" of what is produced with acquired ability. Second language learners, however, have another "mode of production" open to them, by substituting their native language to generate utterances in conjunction with a heavy use of learned L2[2] (in other words, by translation). "Monitor underuser" is Krashen's term for glib language performers with poor control of the grammar, and "monitor overuser" is his term for language performance in the "L1 + Monitor" mode.[3]

(4) *The Input Hypothesis*, which states that "comprehensible input" is the only essential requirement for acquiring language. Following Chomsky's (1965) postulation of an innate human capability to acquire language without any formal instruction at all, this hypothesis holds that people acquire new language forms merely by understanding messages that are encoded at just slightly beyond their present stage on the natural developmental order (Krashen's symbolic expression for this new level is $i + 1$). Language acquirers accomplish this feat of understanding beyond their present level by relying on whatever contextual clues they can find in the discourse, or by "negotiating meaning"[4] until they eventually attain the capability of understanding the new language in decontextualized situations.

In defining what does or does not constitute input, Krashen advances what he calls "the Complexity Argument"[5] : Of all the world's languages, English is one of the best described, having been studied formally for centuries. But even so, the

formal rules of English grammar are so complex that descriptive linguists still have not been able to codify them all. It is therefore quite unreasonable to expect students to learn a language by studying its formally codified aspects. On the other hand, millions of people have acquired English in the past simply by using it, and millions more will no doubt acquire it in the future, likewise without learning many of the formal rules at all. This casts doubt on the need for consciously learned language, while at the same time making a strong case for the indispensability of subconscious acquisition.

(5) *The Affective Filter Hypothesis*, which argues that thestudent must be emotionally open to the language input for it to become acquired. Such factors as pressure, nervousness, or xenophobia are possible inhibitors that filter out potential "comprehensible input," reducing the amount of "intake" and rendering acquisition impossible. Although somewhat of a mixed metaphor, it has become standard in the professional literature to refer to the affective filter as being "up" or "down," "raised" or "lowered."

Criticism of the Monitor Model has centered on the argument that, although Krashen presents it for consideration as a scientific theory, he has formulated it in such a way as to make experimental disconfirmation difficult, if not impossible (Gregg, 1986; McLaughlin, 1987.[6] For the latest collection of such critical essays, see Barash & James, 1994.). Krashen, however, is well aware of the stringent requirements imposed on any proposed scientific theory (Krashen, 1983) and is concerned with the practical use of his theory in the language classroom (Krashen 1983, 1990). For the past fifteen years he has devoted himself to explicating the Monitor Model and its component hypotheses and applying them to practical teaching concerns. Based on the Monitor Model, Krashen & Terrell (1983) proposed a more-or-less complete teaching methodology, which they termed the "Natural Approach". He has also considered its implications specifically in regards to writing instruction (Krashen, 1984), spelling instruction (Polak & Krashen, 1988), and reading instruction (Krashen, 1993).

Krashen's emphasis has changed over time from the Monitor Model as a whole to an increasingly sharp focus on the Input Hypothesis itself.[7] He unequivocally stated his commitment to the Input Hypothesis in Krashen (1980): In that paper, he clearly relegated the other four hypotheses to subordinate status, serving as preliminaries to, or modifiers of, the Input Hypothesis, with the pronouncement: "In my opinion, the Input Hypothesis may be the single most important concept in second language acquisition today." His publication output since then has never

wavered off the theme of "comprehensible input," sometimes abbreviated as CI.

> As is well known by now, the Input Hypothesis forms part of what I call, perhaps audaciously, a theory of second-language acquisition, and it has become clearer to me over the last few years that the Input Hypothesis is the most important part of the theory.
>
> The Input Hypothesis claims that we acquire language in an amazingly simple way — when we understand messages. We have tried everything else — learning grammar rules, memorizing vocabulary, using expensive machinery, forms of group therapy, etc. What has escaped us all these years, however, is the one essential ingredient: comprehensible input.[8]
>
> ....
>
> ...My conclusions [are] that the Input Hypothesis, in conjunction with the general theory it is part of, is easily able to defend itself against... challenges and makes very plausible predictions about a variety of interesting and as yet poorly investigated phenomena in second-language acquisition. .... My view is that we have scarcely begun to apply it — to the extent it is applied, to that extent will our language programmes be more productive and efficient for our students and easier and more pleasant for teachers.[9]

Krashen (1991) updated the Input Hypothesis to account for five other supplementary or competing hypotheses (either to support them or to refute them). The five are: (1) The Reading Hypothesis, a "special case" of the Input Hypothesis; (2) The Simple Output Hypothesis, abbreviated as SO; (3) the Skill-Building Hypothesis, (SB); (4) The Output Plus Correction Hypothesis (OC); and (5) The Comprehensible Output Hypothesis (CO).

> Methods are often combinations of the hypotheses. Natural Approach rests on the Input Hypothesis, but allows some SB and OC.... Traditional form-based language teaching is usually a combination of SB and OC, with some CI (sometimes included inadvertently). "Communicative" language teaching seems to be a combination of CO and CI. "Whole language" seems to be a combination of CI,SO, and CO.
>
> ....I argue that only the Input Hypothesis is successful in accounting for the data in language acquisition and in the development of literacy. Specifically:
>
> (1) Only comprehensible input is consistently effective in increasing proficiency; more skill-building, more correction, and more output do not consistently result in greater proficiency.
>
> (2) Methods with more comprehensible input consistently win in method comparison research.
>
> (3) Output and error correction do not exist in great quantity. Thus, strong versions of hypotheses that depend on output and correction cannot be correct.
>
> (4) Clear gains and even high levels of proficiency can take place without output, skill-building, error correction, and comprehensible output. .... High levels of proficiency cannot take place without comprehensible input.[10]

In applying the Input Hypothesis to the question of writing instruction (1984, 1985, 1993), Krashen has been forced to rely largely on studies of English-speaking children learning to write in their own native language, as few studies have concentrated on the writing abilities of foreign language students. Nevertheless he confidently asserts that learning to write well entails only two factors: "acquisition of the code" of written English, and "an efficient composing process" (Krashen, 1984). In Krashen's view, once a student has acquired some of this code, formal instruction can help to shape and direct the written output — but the acquisition itself can only be achieved through comprehensible reading input. Applications of the Input Hypothesis specifically in terms of reading are discussed in Krashen (1989) and (1993), where he specifically argues that self-selected extensive reading (in his terminology, *Voluntary Free Reading*, or *VFR*) canbe among the most valuable sources of comprehensible language input, for second language learners as well as for first language learners.

Returning again to the "Complexity Argument," Krashen credits Voluntary Free Reading with an almost miraculous power to improve students' spelling, grammar, and vocabulary more effectively than overt formal instruction can:

> .... As is the case with oral language acquisition, competence in writing does not come from the study of form directly — the rules that describe written language...are simply too complex and numerous to be explicitly taught and consciously learned. We gain competence in writing the same way we gain competence in oral language; by understanding messages encoded in written language, by reading for meaning. In this way, we gain a subconscious "feel" for written language, we acquire this code as a second dialect.
>
> This hypothesis accounts for the research showing a relationship between pleasure reading and writing ability — better writers have acquired the written dialect via reading. It also accounts for the lack of a relationship shown between grammar study and writing ability....[11]

Taken together, Krashen is insisting that we learn to read by reading, not by formally studying how to read. Moreover, by reading we learn to write, too! Writing instruction is helpful *after* the student has acquired the written code *by extensive reading*.[12]

> The theory implies that a central task for any language arts programme is getting students hooked on books. .... The expense of books, even paperbacks,... place [s] an increased burden on the school, on the language arts teacher, and on the school library. The theory and research clearly imply, however, that an investment in encouraging pleasure reading, certainly a modest one when compared to expensive educational technology, will pay off in better writing.[13,14]

EXPERIMENTAL DESIGN, SUBJECTS, METHOD, AND MATERIALS

Testing the effectiveness of such an extensive reading approach on writing ability was precisely the point of my study. The experiment was conducted during the 1992 academic year at Nara University, a private four-year coeducational institution located in Nara, Japan. Fourteen sections of Freshman and Sophomore students served as unwitting subjects. Student writing samples were collected in April and December of 1992, prepared for scoring in the spring of 1993, and scored during that summer. The scores were put into a database program in the winter of 1993-94 and statistically analyzed during the spring and summer of 1994.

My intention was to make this research project as pure as possible, but, due to a flawed experimental design and operational errors of my own making, the project was unexpectedly problematic and underwent considerable modification after the orginal proposal had been made. These operational errors and forced changes of design no doubt contaminated the resulting data, rendering them at best ambiguous and inconclusive.

In contrast to American colleges and universities, where students in good standing may freely choose their majors and may change majors at any time, Nara University, like virtually every other Japanese tertiary institution, admits students to a particular department, not to the university as a whole. Within each department, furthermore, different categories of students are admitted under different criteria. For each of Nara University's two upper-division departments there are separate sets of internally-generated entrance examinations for two categories of students, "regular" and "recommended". An English test is one component of each of these these four sets of examinations. Not only are these examinations not standardized (for reasons of test security), but still other candidates are admitted into Nara University without any English examination at all, In a "specially recommended" status. Thus, it is very difficult to get a reliable assessment of the English level of Nara University students in general, or of each department's students, or of each major's students.

No English linguistics or English literature major is offered at Nara University and few students express a strong desire to develop their English proficiency, having been attracted to this particular school for the pursuit of other goals. As part of their general education requirements, however, all Freshmen and Sophomores are enrolled in two English classes per week, each conducted by a different instructor. Students are not streamed by proficiency, nor do they have any individual choice in their section assignments. Instead, they are assigned to sections

according to major and student identification number (corresponding to alphabetical order by surname). For five of the majors, each section accounts for about ⅛ of the population; for the sixth major, each class is half of the total population. We might assume that each section accurately represents the whole population of that major, but a wide variance in proficiency would make such an assumption untenable, unless each section were sufficiently large. Class sizes in 1992 ranged from about 35 to almost 60, depending on the specific major. The statistical comparison of sub-groups is a tenuous undertaking, then, and conclusions must be viewed skeptically.

Until I arrived on the permanent faculty in 1992, Freshman and Sophomore general education English classes were all taught by Japanese staff members. It was not my intention to revolutionize the curriculum nor upset the *status quo* in any way. Like the other permanent English teachers, I was assigned seven sections of Freshman and Sophomore English, so I chose textbooks similar to the ones my Japanese colleagues were using and planned to conduct my classes as similarly as I could. Very soon into the academic year, however, I realized that I was not capable of teaching English in the way that Nara students expected, as a translation-oriented intensive reading course. At that time I abandoned all hope of teaching with a methodology that myself could not deliver, deciding instead to conduct class in my own way. I essentially abandoned the textbooks I had selected and hurriedly threw together an "extensive reading" program, based on my own personal sources of reading materials, accumulated over several years, which I carried by the bagful into each class session and encouraged the students to try.

Krashen (1992, 1993) specifically and pointedly recommends comic books as extremely valuable sources of language input. I had no comic books at my disposal, but I did have my own children's seven-year collection of American magazines to draw upon.[15] Beginning with a wide variety of one- or two- page fiction and non-fiction pieces, I encouraged my students always to choose an interesting topic written at a level they could understand without strain. Gradually I introduced whatever books I could scrape together, ranging from graded readers to books for native English-speaking juveniles to books for the native English adult audience.

It was essentially a leap of faith: Although I had often used some reading materials as a component of conversation and composition classes at my previous appointment, I had never taught a reading class *per se* before and had certainly never based an entire course on the materials I had collected. However, my desperate situation seemed to have provided me with an opportunity to do an extremely pure statistical research study with large numbers of subjects, so I used my circumstance

as the basis for this project.

Originally, I proposed a study comparing the effects of different styles of English courses on the English writing ability of two groups of about 300 students each. The members of one colleague's seven ordinary sections of Nara University Freshman and Sophomore English were to comprise the control group. The members of my own seven Freshman and Sophomore sections were to become the experimental group. One of the experimental group's two English classes per week was to continue as usual, the other weekly class (my class) was to be devoted entirely to extensive reading. For both groups, the "other" teacher each week was a third party, not involved in the conduct of this study, and therefore an uncontrolled factor. Robb & Susser (1989) report a similar consideration possibly contaminating their otherwise well-controlled reading project.

Following the same procedures I had developed over several years at my previous appointment, my Nara colleague and I elicited 30-minute writing samples from both groups (all 14 sections). Each subject was given blank paper and a photocopy of an untitled, four-frame set of wordless picture-cues (which I will refer to as "The Purse-Snatcher") from Byrne (1981), and instructed to write as complete a story as possible in 30 minutes. Dictionary use was not restricted, on the grounds that the accuracy advantage of dictionary consultation would be counterbalanced by a decline in the fluency rating, due to the time consumed. Besides, without dictionaries, students at this level of proficiency could not be expected to produce much at all, even in the "L1 + monitor" mode.

My intention was to provide the baseline data for a gains study of a simple pre-test/post-test design, comparing the writing ability of the control group, taught English entirely in the customary way, against the writing ability of the experimental group, who would have extensive reading substituted for half of their instruction over one academic year. The same task given again in December was to be the post-test, and the comparison between the two samples would constitute the overall gain. "Teaching to the test" was not a feature of either group's instruction, as neither group had a composition class of any kind and "The Purse Snatcher" had no obvious connection to the class material for either group.

Stating the experimental hypothesis in testable terms: On the basis of three objective criteria, Nara University Freshman and Sophomore students enrolled in an extensive reading class will make gains in English writing proficiency equal to greater than comparable students enrolled in traditional English skills classes.

To keep the comparisons truly pure, I wanted to make the extensive reading

classes exactly that: reading only, with no instruction in grammar or vocabulary, and with as little required output as absolutely necessary. I requested that the students try to read 1,000 words per day (and they all groaned). The comparison between the experimental and control groups was thus to be a measure of the success rate of classes of overt formal study against classes of loosely-supervised, non-interventionist extensive reading, insofar as possible conducted in the spirit of Krashen's (1992) observation:

> In my interpretation of the research literature, one kind of reading appears to be the most effective: Free voluntary reading — reading because you want to, with no book report and no obligation to finish the book you started.[16]

So, after a few weeks of preliminary instruction on reading skills techniques, I taught the students virtually nothing for the remaining sessions of the first semester. Instead, I used the class time for free reading and one-on-one personal conferences about the quality and quantity of their free reading.[17]

Other than exhorting the students to be honest to themselves, I made no attempt to exert any control over their choices. Mainly, my purpose in conducting the in-class personal conferences was to encourage the students not to dissipate their time and energy in trying to read books of too high a level, but instead to choose without embarrassment material at the best level for themselves—even simple books—that they could honestly understand and would enjoy reading. I asked them to try to read at least 1,000 words per day, every day, including weekends and holidays. If the material they had chosen turned out to be dull or too difficult, I encouraged them to exchange it for something "better" rather than struggle on with material that was "no good" for them.

From the perspectives of four of the five Monitor Model hypotheses, this advice was essential, but I have no doubt that the very concept of deriving *enjoyment* from the reading of an English book was unthinkable to most of the students. Yet, on an attitude questionnaire administered at the end of the academic year, many students responded that 1,000 words per day was not too much to ask, and several students reported enjoyment. For some of these students this free-reading opportunity may have been the first time to have ever felt successful at anything connected with English. A colleague also reported one of my students boasting to him about how many English books he had read so far that year. Near the end of the year one student, a bicycling enthusiast, told me that he was glad to have learned how to read equipment reviews in English-language cycling magazines.

Since each student was reading selections of his or her own choosing, it was

impossible for each student's comprehension to be checked, in Casanave's sense — the teacher ensuring that the students understood what they had read (Casanave, 1988). The students were required, however, to keep records of their readings using an estimation method, honor-bound to report their daily reading correctly. They were asked to include cumulative totals of their daily amounts, along with their own subjective assessment of each day's material in terms of interest (YES or NO) and level of difficulty (EASY, OK, or HARD), and to graph the daily fluctuations in their reading speed (in WPM). On a year-end anonymous "honesty" questionnaire, some of the students admitted violating this trust blatantly, but most of the students claimed to have reported their reading inside tolerances consistent with the estimation method that I had encouraged.

Although most students seemed to have conscientiously read English as homework, few students utilized their unsupervised class time for English reading and the first semester's class hours were essentially wasted, except for those few students each time with whom I was able to confer briefly. My grandly-conceived experiment had thus quickly evolved into a comparison of traditionally-conducted *classwork* against instruction-free classes of extensive reading *homework*.

In what I now consider a major procedural error, I had the idea after summer vacation that taking a mid-year writing sample would allow me to examine the possiblity of a "practice effect" — that is, the extent to which the mere experience of writing the story could affect subsequent trials. I had already sampled four of the seven experimental group sections in September before my colleague reported the shocking news that most of the control group subjects had rebelled at providing a mid-term writing sample. Knowing that my results would be compromised if there really were a pronounced practice effect, I immediately cancelled sampling my own remaining three sections, so that in December at least some members of the experimental group could be comparable to the control group. What had begun with about 300 subjects in each group was thus now reduced in one blow to only about half that number in the experimental group.

I altered the conduct of my class for the second semester, mainly because I feared for my professional reputation. Regardless of how persuasive Krashen's hypotheses may be to me, they are still only hypotheses, and I began to feel that if I were criticized for having "taught" the students virtually nothing in one year, I could not have justified my actions. For that (admittedly selfish) reason, I decided to

assign some skill-building classwork in the second semester, to occupy the students while I was conducting personal conferences. For this purpose I used handouts of manuscript pages from an unpublished reading skills textbook (Day *et al*, in press). A brief amount of time was spent in presenting the skills lessons at the beginning of each class session and in reviewing the answers together at the end. I neither checked the students' work nor graded them, but to the extent that the students exerted themselves on the lessons in class, they received skill-building instruction, not just simple extensive reading, and the second semester therefore constituted a somewhat less pure test of the Input Hypothesis than the first semester had been.

I collected the December writing sample from the experimental group sections without any serious incident. My worst fear was realized, however, when, despite my colleague's efforts to encourage her control group students to participate in the post-test sampling, most of them refused to co-operate, leaving me no means of comparison with the experimental group. I was stuck, then, with seven Control Group sections that had been sampled only once, three Experimental Group sections that had been sampled only twice, and four other Experimental Group sections that had been sampled three times. Worse, at mid-year I had sampled both of my two Sophomore sections, making comparisons between the two Sophomore sections impossible.

My plan to use a Nara control group was spoiled, but before completely abandoning the comparative gains design, I tried comparing the Nara experimental group's results against data I had collected over several years at my previous appointment, Baika Women's College, a respectable four-year college in Osaka, where I had taught only English literature majors. My hope in this desperate last-resort tactic was that extensive reading alone would enable the Nara experimental subjects to show comparable gains in writing proficiency, despite the Baika students having had every other possible advantage (English majors, relatively higher interest in English language and literature, twice as many contact hours per week in smaller classes, specifically including composition classes, extracurricular English Speaking Society, and school-sponsored opportunities for travel to English-speaking countries).

At Baika I had collected similar student narrative writing samples for several years and had written (Swan, 1987) about the difficulties in applying what had seemed to be a promising holistic rating scale (Jacobs *et al*, 1981), but which had ultimately proved disappointing. This time I used a more objective rating technique. First, as a coarse measure of fluency, I simply counted the number of words in each writing sample. The other two objective criteria used for this study are those identified by

Ishikawa (in press) as the two best predictors of language proficiency: (1) the number of words appearing in perfectly-formed clauses; and (2) the number of perfectly-formed clauses in the writing sample. A perfectly-formed clause is operationally defined by a set of protocols Ishikawa created to help ensure rater reliability — that is, ensuring comparable scores from different raters or consistent scores from the same rater over time. As objective as Ishikawa's method is, however, there is still some latitude for the raters to have individual grammatical interpretations.

Six college EFL teachers, all of them Americans with years of experience in Japan, were trained by Ishikawa and served as raters for this project. Five of the raters each scored the pre- and post- writing samples of comparable numbers of Nara and Baika Freshman subjects. The sixth rater had no Baika data to evaluate, but scored the samples from both of the Nara Sophomore sections. As a safeguard against rater unreliability, I had hoped to have each rater's scores double-checked by another rater, but in practical terms a double-checking procedure was not feasible, as it would have doubled each rater's load and would have been simply too much work to ask of volunteers.

The ratings for each of the three criteria were analyzed by the paired t-test procedure. Comparisons between the Baika Freshmen and the Nara students were fruitless, however. Starting from incomparably high levels of English ability, the Baika students went on to outdistance the Nara students embarrassingly. The difference in mean scores was simply too great to permit the valid use of Baika students as a control group.

In light of all these design modifications, procedural errors and failures, the only avenue left for me was to revise the experimental hypothesis and examine the Nara University experimental group without reference to any control group, making the experiment no longer a comparative gains study at all. The new hypothesis reads: On the basis of three objective criteria, Nara University Freshman and Sophomore students enrolled in extensive reading classes will make significant gains in English writing proficiency.

Neither the Nara control group nor the prospective Baika control subjects are discussed any further in this paper.

RESULTS

Table 1 presents the results of paired t-tests performed on the gains made by the aggregate Nara experimental group (n=295).

	FLUENCY	ACCURACY	
	Number of words written in 30 minutes	Number of words appearing in perfect clauses	Number of perfect clauses
PRE-TEST			
t-value	70.5186	20.3424	4.3288
Standard Deviation	30.817	6.409	3.480
significance	***	***	***
POST-TEST			
t-value	98.0814	24.7153	5.5898
Standard Deviation	40.375	19.960	4.318

TABLE 1 : Comparison of Pre- and Post- Writing Samples, Aggregate Experimental Group, Rated for Fluency and Accuracy

n.s.=not significant *p<0.10 **p<0.05 ***p<0.01

The t-tests show highly significant gains on each of the three criteria. The confidence levels of these gains are so high that they can not be computed within three decimal places, meaning that the likelihood of this average gain occurring by chance is vanishingly small, less than one chance in 1,000. It would seem that the null hypothesis is soundly rejected, and that extensive reading did indeed improve the students' writing ability. There is no way to know how much the Nara control group may have improved, however, since insufficient post-test data was collected from the control group students.

Since there was no control group to compare these gains against, I thought it might be interesting to see how well the Freshmen performed in comparison to the Sophomores. As judged from entrance examination results, however, students from the six different majors have historically had different average levels in English ability. It seemed only fair to eliminate the Freshman history majors from consideration,

then, since no Sophomore history majors had been sampled. This left only five sections of experimental group subjects: two sections of Freshman geography majors (one sampled at mid-year, one not sampled); one section of Freshman Cultural Studies majors (not sampled at mid-year); one section of Sophomore geography majors (sampled at mid-year); and, finally, one section of Sophomore Cultural Studies majors (sampled at mid-year). The fact that each section was scored by a different rater is a possible source of measurement error, as of course is the fact that not every class was sampled at mid-year.

Even after removing the history majors, t-tests of the remaining aggregate data still show similarly high statistical confidence levels, incomputable within three decimal places, for the gains over one year. The results of the paired t-tests on this subset (n=194) are presented in Table 2. Table 3 presents the results of paired t-tests on the subset sorted by year (Freshman n=120; Sophomore n=74).

	FLUENCY	ACCURACY	
	Number of words written in 30 minutes	Number of words appearing in perfect clauses	Number of perfect clauses
PRE-TEST			
t-value	70.5309	17.6959	3.9536
Standard Deviation	32.084	15.709	3.397
significance	***	***	***
POST-TEST			
t-value	95.9897	20.8505	4.9021
Standard Deviation	39.418	16.825	3.827

TABLE 2: Comparison of Pre- and Post- Writing Samples, Experimental Group Subset (101 Freshman history majors removed), Rated for Fluency and Accuracy

n.s.=not significant *p<0.10 **p<0.05 ***p<0.01

	FLUENCY		ACCURACY			
	Number of words written in 30 minutes		Number of words appearing in perfect clauses		Number of perfect clauses	
	PRE-	POST-	PRE-	POST-	PRE-	POST-
FRESHMEN						
t-value	75.6250	92.0167	19.1667	18.8000	4.1750	4.2500
Standard Deviation	34.127	37.360	17.010	14.969	3.361	3.168
significance	***	n.s.	n.s.	n.s.	n.s.	n.s.
SOPHOMORES						
t-value	62.2703	102.4324	15.3108	24.1757	3.5946	5.9595
Standard Deviation	26.668	42.007	13.043	19.105	2.965	4.531

TABLE 3: Comparison of Pre- and Post- Writing Samples,
Experimental Group Subset (101 Freshman history majors removed)
Sorted by Year, Rated for Fluency and Accuracy

n.s.=not significant *p<0.10 **p<0.05 ***p<0.01

The Sophomores and the remaining Freshmen both show gains on each criterion, except the "number of words appearing in perfect clauses" criterion, for which the Freshmen show a very slight average decline.

Examining the scores derived from the pre-test writing samples, we see that the Sophomores' April mean scores were lower than those of the Freshmen for all measures, though only the fluency measure is statistically significant, at the confidence level of .005. On that measure, the Freshmen outscored the Sophomores in April by an average of more than 13 words, 75.6250 against 62.2703. For the two accuracy measures, the Freshmen tended to outperform the Sophomores in April, but at statistically insignificant levels of confidence.

On the December post-test, however, the situation was reversed. Although the post-test differences between the two years are statistically insignificant, the Sophomores not only closed the gap, they went on to surpass the Freshmen on all three criteria. Both groups showed gains in the fluency measure, but the gain was greater for the Sophomores than for the Freshmen: Besides having started out at a

lower level, as previously noted, the Sophomores ended up at a higher level, 102.4324 (△40.2), compared to 92.0167 (△16.4) for the Freshmen. On the post-test "number of words appearing in perfect clauses" criterion, the Sophomores outperformed the Freshmen again, 24.17 (△8.86) vs. the Freshmen's average of 18.80, a drop from their April score (△−.3.667). On the post-test "number of perfect clauses" criterion, the Sophomores outperformed the Freshmen, 5.95 (△2.4) vs. 4.25 (△.08).

Why should the Sophomores have improved so dramatically in comparison to the Freshmen?

To examine whether there was some sort of practice effect induced by the sampling itself, I separated those Freshman sections which had been sampled in September (n=91) from those which had not been (n=86). Fortunately, I had sampled only one each of my two history sections and two geography sections, leaving one comparable section of each major unsampled. Sorting them by frequency of sampling, I then analyzed the yearlong changes for each criterion. The results are presented in Table 4.

	FLUENCY	ACCURACY	
	Number of words written in 30 minutes PRE-POST difference	Number of words appearing in perfect clauses PRE-POST difference	Number of perfect clauses PRE-POST difference
NOT SAMPLED AT MID-YEAR			
t−value	15.8372	−.7558	.1860
Standard Deviation	32.873	5.611	2.892
significance	***	***	***
SAMPLED AT MID-YEAR			
t−value	31.9670	8.1099	1.9780
Standard Deviation	4.554	17.933	4.112

TABLE 4: Comparison of Pre- and Post- Writing Samples,
Comparable Freshmen Subset Sorted by Frequency of Sampling

n.s.=not significant *p<0.10 **p<0.05 ***p<0.01

For each criterion, the sections which had been sampled in September (n=91) outperformed the sections which had not been sampled (n=86), at very high levels of statistical confidence. My overall results, showing general improvement in the aggregate data, are therefore suspect, due to the possible existence of a practice effect. Since 100% of the Sophomores were sampled in September but only 41% of the Freshmen were, the existence of a practice effect might be one possible explanation for the surprising turnabout in mean scores between the Freshmen and the Sophomores.

DISCUSSION

Whether we can draw any valid conclusions based on the data from this study is open to question. For the sake of discussion, however, let us follow where the data lead us. These ambiguous results may no doubt be interpreted in many ways; here is mine:

Knowing the circumstances that language instruction must operate under in Japan is the first step needed before we can make any statements at all.

Consider our experiences with the English language competence of the typical Japanese university-age student: He or she has studied English for six years in junior and senior high school, generally without much interest but instead from motivations such as compulsory education requirements, concern for future livelihood, or avoidance of teacher/parent disapproval. As is typical of an EFL (English as a Foreign Language) situation, the opportunities for the typical Japanese student to meet and interact in a natural English setting are extremely limited. In such a curriculum environment, translation-based teaching methods eliminate the need for teacher and student to "negotiate meaning". Instruction is given according to a curriculum and pace determined by the central grovernment, with artificially imposed requirements to present a specified volume of material in the prescribed length of time, regardless of each individual student's ability to internalize it at that pace. Students who are unable to keep up in school classes are forced either to enroll in expensive after-school private classes or fall hopelessly behind the standard, losing forever the chance to enter any university which sets a high standard of English proficiency as one of its obstacles to admission. University entrance exams, upon which the students' very futures are at stake, are created school by school, department by department, not necessarily for reliability or validity but with a great concern for administrative convenience. In many cases, they do not ask the student

to communicate in English, but to solve linguistic puzzles, such as choosing the best answer from a menu of sometimes pointlessly subtle nuances, modified cloze tests, usually of prepositions or articles (the two most inscrutable categories and the last to be mastered), paper tests of accent, intonation, or pronunciation (probably the most artificial language environment imaginable), or translation passages — either of arcane literary works or of highly technical essays in the social or physical sciences.

The result of this set of circumstances is that, after six years of formal study at a very advanced level, many Japanese students do, in fact, have some degree of English reading ability; given dictionary access and enough time, a great many of them are able to puzzle out whatever English sentences that may be placed in front of them. The ability to read with comprehension in "real" time, however, is lacking in most high school graduates.

English language production is another matter, too, whether in speech or in writing. Most students depend on laboriously memorized grammar rules and vocabulary lists (or heavy dictionary consultation) to generate even the simplest of English utterances—in any and all situations. Virtually every response requires an inordinate amount of time and effort (and listener patience), due to a lack of spontaneous productive ability, and a consequent, agonizingly slow, mental search through the maze of memorized rules.

Nevertheless (in confirmation of the "Complexity Argument"), even with all this laborious process of data-retrieval, it is a rare student indeed who, after six years of secondary-level language study, can produce an error-free English sentence beyond the simplest of formulaic greetings.

Though Krashen's theory may be criticized on formal grounds, its descriptive power is still attractive. Rephrased in Krashen's terms, the typical Japanese student's six years of laborious study is language *learning*, at a level of difficulty far beyond his or her level of language *acquisition* along the *natural order*. This learning has usually taken place in large classes of 30 or more students, where, in almost complete disregard of *affective* considerations and in contravention of the "Complexty Argument," grammatical correctness on translation examinations is valued above spontaneity of expression in "real" time. These factors virtually require that classes be conducted without much contextualization to help make the new language directly comprehensible—in other words, conducted by translation methods. To survive the prescribed pace of instruction, students learn to *overmonitor* themselves at all times. In short, little of the students' six years of English

exposure qualifies as English *input*, in Krashen's sense. Although much is "taught" and even "learned" in six years, for the vast majority of students little or nothing is *acquired*. Some students thrive in such an environment, but most do not. Sadly, these students generally believe that the fault lies not with the system under which they must operate, but with themselves, for not having been bright enough or not having studied diligently enough.

In my two decades of experience in Japan, I have noted a consensus opinion that that the typical university student's English proficiency begins to decline soon after university entrance examination are finished and the pressure is off. Since they have come to depend on high levels of *learned* language rather than natural levels of *acquired* language, by the time the students begin classes in April, many of them are already several months into the process of forgetting everything they ever knew about English, and (except for those who become English language majors or those who have some other vested interest) this decline has generally been observed to continue throughout their university careers.

The foregoing description of the existing Japanese situation is not to be taken as a criticism of Japanese students, or even of Japanese teachers, for that matter. Similar conditions and results surely pertain in other parts of the world where the language to be learned is not the language of the country. It is the difference between an FL (Foreign Language) situation and an SL (Second Language) situation, where the language to be learned is the language of the surrounding society, not only existing within the classroom.

It is natural and reasonable that foreign language production ability lags receptive ability, but among most Japanese students the gap is simply too wide. Rectifying this disparity is the central problem that language educators in Japan must address. In response to this problem, Krashen would say that what is needed is not more laborious study, leading to ineffectual language *learning*, but more comprehensible input, leading to natural language *acquisition*.

What statements can we make from the results of this study, then?

To begin with, meaningful language education research is hampered by the failure of private Japanese institutions as a whole, and Nara University in particular, to gather standardized language proficiency data for admissions purposes. This is a given circumstance under the present system of university admissions in Japan, and, unless the entire country's education system were changed, this research handicap will continue. There are many factors, not the least including economic ones,[18] discouraging such change within the foreseeable future.

If some kind of standardized proficiency score were available, however, we would predict large variances in each section at Nara University at the beginning of each academic year, due to the students of each major having been admitted to the university under varying criteria then simply lumped together at the same level of English instruction. For the purposes of this study, the pre-test writing elicitation must serve as the nearest thing to a standardized baseline test, and a glance at the numbers available here supports the above prediction of wide variance (for example, the Freshman fluency mean of 75.6250 with a standard deviation of 34.127, or an accuracy mean of 19.1667 with a standard deviation almost as wide as the mean itself, 17.010; meanwhile among the Sophomores a fluency mean of 62.2703 with a standard deviation of 26.668 and an accuracy mean of 15.3108 with a standard deviation of 13.093).

Corresponding to the wide variance in proficiency levels generally, we would also expect the section-by-section variances within the same majors to be wide, since each section comprises ⅛ of the total within each major and since the proficiency distribution is determined randomly by student ID number.

The administrative policy decision not to stream English students by proficiency levels, while fostering cooperative social attitudes that the Japanese consider important, renders meaningful language teaching nearly impossible by traditional classroom methods. The reason is neatly explained by Krashen's concept of $i + 1$: Regardless of what level of formal classroom instruction any English instructor aims at, it is virtually guaranteed to be *inappropriate for the majority* of the class, at best st being just right for only a small band of students somewhere within the vast distribution of proficies.

This is one of the reasons that a non-interventionist, self-selected extensive reading approach holds such appeal: it allows each student the freedom to seek language input at his or her own best level. Conducting class as I did could be considered professionally irresponsible, perhaps even outright unethical, if Krahen's data and arguments in support of extensive reading programs were not so persuasive, if admission and administrative policies were not already so stacked against the students' success, and if the results of traditional interventionist teaching methodologies were not already so disappointing.

Under the assumption (not necessarily a valid one) that each subset is representative of the larger population, the difference between the pre-test scores of the Freshmen and Sophomores may be interpreted in three ways: Since every Freshman section was scored by a different rater, one possibility is interrater unreliability－the

concept that, for some reason, the Freshmen's five raters scored their subjects' writing samples at a quite different stringency level than the Sophomore's solitary rater did. We can not be really sure to what extent this may have occurred, but it is so improbable at the observed levels of statistical significance that it may be safely ruled out. A second interpretation of the gap between Freshman and Sophomore scores could be that candidates of significantly different English proficiency are admitted in different years, the consequence of administering non-standardized examinations and of maintaining differing standards for various segments of candidates. The third possible interpretation of the gap is that one year of instruction at Nara University resulted in the Sophomores' net loss of English proficiency from the time of their admission. Although we have no data for them as Freshmen, this view corresponds with the many years of past observation previously noted. None of these interpretations can be verified directly, but it is reasonable to assume that there is some truth to all three: (1) This data may be contaminated to some degree by operational error or rater unreliability but, beyond that, (2) because of the lack of standardization on the entrance exams, we can never be quite sure of the relative English proficiency of each year's entering Freshmen, though (3) we can be fairly sure that their average performance will likely decline once they have passed through the pressure of entrance exams and have begun university.

Over the year there was a general rise in the writing performance of the aggregate group on every measure, which, in the face of our previous expectation of a general decline in proficiency, I take as impressive empirical evidence in support of the Input Hypothesis. On every measure, the aggregate Sophomores began the year with a lower average than the Freshmen, but ended the year with a higher average. Regardless of whether the Sophomore pre-test scores are considered as an indication of language attrition during their Freshman year or as an indication of their prior inferiority to the current Freshmen, the reversal implied by their post-test ratings is stunning and can hardly be attributed to rater unreliability.[19] It is very difficult to maintain the proper degree of academic objectivity and caution in light of such one-sided figures. At the same time, it is a complication for which I have no persuasive explanation.

Inconveniently, the existence of a writing practice effect also seems to be suggested by the data, or at least can not be ruled out. Krashen might attribute results such as these to the improvement of the students' "composing process," but if there were indeed a practice effect, its existence could also be construed as empirical support for one of the contradictory "Output" hypotheses — that we learn

to speak by practicing speaking and to write by practicing writing, not necessarily by receiving comprehensible input.

Another complication is that one Sophomore class vastly outperformed the other (data not shown). It may simply be that one section more eagerly seized the opportunity for freedom offered by the extensive reading class, but it is also possible that below a certain threshold of proficiency this approach is less effective, at least within the difficulty levels of the reading materials at my disposal. With so few students being compared, it is difficult to make any firm statement on such matters. For all we know, it might even have been caused by the time of day the classes were held or the temperature in the classroom during the taking of the writing sample — a dozen possibilities suggest themselves.

CONCLUSIONS

Even without the unforeseen rebellion by the control group and the logistical impossibility of providing rater reliability checks, the observed data would have probably been compromised from the beginning by the poor research design and clumsy execution of this study. All conclusions must therefore be viewed as tentative, until confirmed or refuted by a more carefully controlled replication.

Nevertheless, despite the extensive reading treatment having been thrown together at the last moment, the results seem to have been highly successful. From nearly fifteen years of past observation, not only at Nara University but at several other institutions as well, I would predict a decline in the subjects' English proficiency after the pressure of entrance exams is off and the students have entered the university. Actually, however, the improvement of the aggregate experimental group was statistically highly significant. We cannot know how well the control group would have done, but the data as observed reject the null hypothesis and support the revised experimental hypothesis: Although the aggregate gain scores tend to break down when subsets are examined individually, the English writing proficiency of Nara University students enrolled in a one-year program of extensive reading improved at statistically significant levels, as measured by the three objective criteria of fluency and accuracy. There is evidence, however, of a possible practice effect, which tends to undercut the support.

In reference to native-language students of English, Krashen cautions that the results of an extensive reading program may be only slowly cumulative, that they may take a long time to be detectable in output performance:

> It may not, however, pay off immediately. .... One doesn't simply assign a few articles or even books and see immediate results — just as is the case in second language acquisition, large amounts of comprehensible input may be necessary before acquisition becomes evident. Thus, the teacher who succeeds in getting a student hooked on books may not see the results; next year's teacher may get the credit.[20]

How much more so, then, for foreign language students, who have much less experience with the language?

Until now there has not been a measure of writing performance sufficiently sensitive to register the changes that might occur in only one year. I was fortunate to have Ishikawa's rating technique available to me for this study, however. In my opinion, it is an extraordinarily useful measurement tool, because it not only allows for unlimited student development at the upper end but also reveals minute but statistically significant increments of improvement at all levels of proficiency. My vain hope that the Baika data could be used as a control for the Nara data was based on my earlier rating experience with the Jacobs *et al* holistic scale. Using Using Ishikawa's measures, however, the difference between the two groups was made painfully obvious. In this trial there were no checks for interrater reliability, but the economic limitations on my project in no way diminish the value of Ishikawa's measurement technique, in my opinion. For applied linguists it is the equivalent of a microscope: It offers us the potential to observe the language acquisition process with an unprecedented degree of precision and, if widely accepted, could lead us to entirely new insights.

SUGGESTIONS FOR FUTURE RESEARCH

In light of the errors I made in designing and executing this study, if I were to try it again, I would want to do several things differently.

A major mistake in my original proposal is that I had planned to conduct all the classes in the experimental group myself and have a colleague conduct all the classes of the control group. In fairness to my colleague, I must note that I made this research design from my own imperfect understanding of the prerequisites for the statistical procedures involved. I only later learned that this design would have invalidated the results even without the rebellion of control subjects, since the analytical procedures could not have been validly conducted on the two intact groups that my design had created. In a future retrial, I would ask more teachers to participate as instructors of both groups.

I would also try to pay much more careful attention to controlling the extraneous

variables. In my desire to have the students explore the English language freely at their own best levels, I allowed the class sessions to be too loosely controlled. While there is certainly an argument to be made that the very concept of voluntary free reading also implies the student's right not to read anything at all, such a degree of freedom is not helpful to the student with limited opportunities for contact with the target language. Since this experiment was performed, I think I have learned how to exert better teacher control over an extensive reading class without confining the students too tightly. My experience comes too late to salvage this trial, however.

Thirdly, I would not complicate matters by simultaneously looking for a writing practice effect. It is clear now that my reach exceeded my grasp. Several possible remedies come to mind. The easiest and most economical way to avoid this complication would of course be to refrain from making a mid-term observation. Alternatively, I could have had the subjects write a different story at post-test time, rather than administer "The Purse Snatcher" again. I had rejected that approach, however, on the grounds that I could not be sure the two different stories would be equally easy for the students to tell. Perhaps the best approach, finally, would have been to have been to give the subjects two opportunities to write the story at pre-test time, a "warm-up" or "practice swing," so to speak, then count only the better of the two as the "official" measure of their pre-treatment writing ability. In that way, whatever practice effect might occur would have been accounted for at the very outset. Any other gains would be attributable to the treatment. The main problem I can foresee with that approach is that the student rebellion would likely have occurred in April, rather than in September or December.

Finally, to my way of thinking, it is virtually impossible to conduct any kind of comparative gains study without a reliable standardized proficiency measure. In this study, the pre-test writing sample itself was the yardstick, but in any retrial I would want the whole population to be screened with an appropriate standardized measure first, then have all the subjects divided into large experimental and control groups at comparable levels of initial proficiency. An ideal study would enlarge the scale to include not only Nara University students but also students from other institutions and from a wide range of majors. The cost and logistical difficulties of such a retrial would probably rise to unmanageable proportions, however, and such an experiment could not be conducted without a much higher level of funding and the enthusiastic cooperation of several institutions and many teachers. Alternatively, a smaller sampling of subjects, such as the students at one institution, could be

selected for comparability before being used as representatives of the whole population. At the very least, the subjects selected for this sort of limited study must be students from the same majors and in the same year of study, and in sufficiently large numbers to ensure that the statistical comparisons are valid.

ACKNOWLEDGEMENTS

Funding for this project was a Special Research grant from Nara University, which is here gratefully acknowledged. Baika data was collected over several years, partly funded by research grants from Baika Women's College, which also are gratefully acknowledged.

The many flaws in the study are my own reponsibility, of course, but this project could not have been completed without advice and cooperation from several colleagues. For their helpful suggestions at the earliest stages, I would first like to thank Steve Ross, Beniko Mason, Bernard Susser, and Julian Bamford. For graciously allowing me to intrude on their 1993 summer vacations, I am indebted to Sandra Ishikawa and the six other American colleagues who rated student writing samples with her technique. In alphabetical order, they are: Don Kaduhr, Kathi Kitao, Mary Goebel Noguchi, Steve Vanderbilt, Kathy Yamane, and Bonnie Yoneda. All seven of them have my deepest gratitude for undertaking that distasteful task cheerfully. Finally, special thanks to two of my Nara University colleagues; first, to Hiromi Moriyama of the College of Liberal Arts, for conducting my experiment in the control group sections, and, finally, to Arinori Yosano of the Department of Social Research for his many hours of invaluable help on the inferential statistics; without his generosity, I would no doubt have given up in despair.

NOTES

1. Adapted here from Krashen, 1985.
2. Adapted here from Krashen, 1981.
3. "The L1+Monitor mode of producing utterances allows second language production without acquired competence. .... This happens especially when performers need to produce utterances very early in their second language experience, before they have had a chance to acquire the needed L2 structures "naturally" via input. Performers simply utilize the surface structure of their first language, and then employ the conscious grammar as a Monitor to make alterations to bring the L1 surface structure into conformity with their idea of the surface structure of the second language." (Krashen, 1982, p. 210.)
4. Krashen, 1985, on Bilingual Education programs is obliquely relevant here: "Ineffective bilingual programmes use the first language in such a way as to block comprehensible input. This occurs when techniques such as concurrent translation are used, in which a message is conveyed to students in one language and then translated into the other. When this is done, there is no need to 'negotiate meaning'; the child does not have to listen to the message in the second language, since he knows it will be repeated in his first language, and the teacher does not have to make an effort to make the English input comprehensible. Research has confirmed this theoretical prediction." (p. 18.)
5. Adapted here from Krashen, 1993, p. 14.
6. There are undoubtedly many more that I am not acquainted with, as evidenced by McLaughlin's remark: "Indeed 'Krashen-bashin' has become a favourite pastime at conferences and in journals dealing with second language research." (McLaughlin, 1987, p. 19.)
7. For a fairly detailed yet concise overview of the development of Krashen's ideas from the 1970s to the present, see Shannon, 1994. For an account of previous work that underlies Krashen's own, see Larsen-Freeman, 1985.
8. Krashen, 1985, vii.
9. Ibid.
10. Krashen, 1991, pp. 409—411.
11. Krashen, 1984, pp, 27—28
12. Krashen, 1984, p.27.
13. Krashen, 1984, pp. 29—30
14. In the two foregoing notes, Krashen refers to LI studies, but his discussion is applied to all situations.
15. Of the several different magazines in my home, the most useful resource for my purposes was *Highlights for Children*. It aims at a relatively wide range of ages and skills, in comparison to other children's magazines, most of which are much more narrowly targeted to a particular age. While some of the the fiction is childish and overly didactic, much of the non-fiction for older children is informative and interesting even to adult readers and may be profitably read by people of all ages.
16. Krashen, 1992, p. 413.
17. I was able to confer with only about seven or eight students per class session, so, since my sections ranged in size from 35 to 58 students, I had barely enough time to meet with each student once per semester. Actually, in one large class of Freshman history majors, I was not able to confer personally with everyone once before summer vacation and had to wind up the last class session conferring with the last contingent in small groups.
18. In contrast to many private schools in the United States, which receive endowments from their successful graduates and charitable foundations, private universities in Japan derive a significant portion of their operating income from the substantial fees charged applicants to take the entrance examination. In the cases of the largest and most prestigious of the private schools, the proceeds from the annual entrance examinations run into the billions of yen. It benefits the administration, the teachers and staff (who collect extra-duty pay for creating and administering the exams), and it obviously benefits the students, but it is similar to a regressive tax on the unsuccessful applicants. Thus does the next generation of losers in Japanese society subsidize the educations of the winners.

19. Both Sophomore sections were rated by the same rater. The data sets were arranged in such a way that the rater would first rate one student's pre-sample and then the same student's post-sample, finishing one section of students before beginning the other section. In this way, a gradual change in stringency would have a negligible impact on each particular student's pre-and post-rating, but would be revealed as a variance between the averages of the two classes. The possibility that the stringency of the rater's interpretations gradually changed over time is a troubling thought. It is impossible to say definitely whether one Sophomore section's April average performances were overrated or the other section's were underrated — or both — but any misrating is extremely unlikely to have occurred at these levels of statistical significance. I believe that the possibility of rater unreliability tainting the Sophomore data may be safely ruled out. Accordingly, the general rise in average performance between April and December is difficult to dispute.
20. Krashen, 1984, p. 30.

REFERENCES

Alatis, J.(Ed.). (1989). *Georgetown University Round Table on Languages and Linguistics 1989 (GURT 89).* Washington, DC: Georgetown University Press.

Alatis, J.(Ed.). (1990). *Georgetown University Round Table on Languages and Linguistics 1990 (GURT 90).* Washington, DC: Georgetown University Press.

Alatis, J.(Ed.). (1991). *Georgetown University Round Table on Languages and Linguistics 1991 (GURT 91).* Washington, DC: Georgetown University Press.

Alatis, J.(Ed.). (1992). *Georgetown University Round Table on Languages and Linguistics 1992 (GURT 92).* Washington, DC: Georgetown University Press.

Alatis, J., Stern, H. & Strevens, P.(Eds.). (1983). *Georgetown University Round Table on Languages and Linguistics 1983 (GURT 83).* Washington, DC: Georgetown University.

Barash, R. & James, C. (Eds.). (1994). Beyond the Monitor model: Comments on current *theory and practice in second language acquisition* Boston: Heinle & Heinle.

Byrne, D. (1981). *Progressive picture compositions.* Longman.

Casanave, C. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly 22:* 2, 283—302.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Day, R., Swan, J. & Yamamoto, M. (in press). *Let's read!* (tentative title). Prentice-Hall Regents.

Gass, S. & Madden, C. (1985). *Input in second language acquisition.* Rowley, MA: Newbury House.

Gregg, K. (1986). Review of Krashen, 1985. *TESOL Quarterly* 20:1, 116—122.

Ishikawa, S. (In press). Objective measurement of low-proficiency EFL narrative writing. To appear in the *Journal of Second Language Writing.*

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V. & Hughey, J. (1981). *Testing ESL composition: A practical approach.* Rowley,MA: Newbury House.

Krashen, S. (1980). The Input Hypothesis. Reprint from Alatis (1980), appearing as the Appendix in J. Oller, *Issues in Language Testing Research.* Rowley, MA: Newbury House.

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon Press.

Krashen, S. (1982). Accounting for child-adult differences in second language rate and attainment. In S. Krashen, R. Scarcella, & M. Long, (Eds.). *Child-adult differences in second language acquisition.* Rowley, MA: Newbury House.

Krashen, S. (1983). Second language acquisition and the preparation of teachers: Toward a rationale. In J. Alatis, H. Stern, & P. Strevens, (Eds.). (1983).. *Georgetown University Roundtable on Languages and Linguistics 1983* (GURT 83). Washington, DC: Georgetown University.

Krashen, S. (1984). *Writing: Research, theory, and applications.* Oxford: Pergamon Institute of English.

Krashen, S. (1985). *The Input Hypothesis; Issues and implications.* Longman.

Krashen, S. (1989). Language-teaching technology: A low-tech view. In J. Alatis, (Ed.), *Georgetown University Round Table on Languages and Linguistics 1989* (GURT 89). Washington, DC: Georgetown University Press.

Krashen, S. (1990). How reading and writing make you smarter, or, how smart people read and write. In J. Alatis, (Ed.), *Georgetown University Round Table on Languages and Linguistics 1990 (GURT 90).* Washington, DC: Georgetown University Press.

Krashen, S. (1991). The input hypothesis: An update. In J. Alatis, (Ed.), *Georgetown University Round Table on Languages and Linguistics 1991* (GURT 91). Washington, DC: Georgetown University Press.

Krashen, S. (1992). Some new evidence for an old hypothesis. In J. Alatis, (Ed.), *Georgetown University Round Table on Languages and Linguistics 1992* (GURT 92). Washington, DC: Georgetown University Press.

Krashen, S. (1993). *The power of reading: Insights from the research.* Englewood, CO: Libraries, Unlimited.

Krashen, S. & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom.* San Francisco: Alemany Press.

Larsen-Freeman, D. (1985). State of the art on input in second language acquisition. In S. Gass, & C. Madden (Eds.), *Input in second language acquistion.* Rowley, MA: Newbury House.

McLaughlin, B. (1987). *Theories of second-language learning.* London: Edward Arnold.

Oller, J. (1983). *Issues in Language Testing Research.* Rowley, MA: Newbury House.

Polak, J. & Krashen, S. (1988). Do we need to teach spelling? The relationship between spelling and voluntary reading among community college ESL students. *TESOL Quarterly* 22: 1, 141-146.

Robb, T. & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language* 5: 2, 239-251.

Shannon, S. M. (1994). Introduction to R. Barash & C. James (Eds.), *Beyond the Monitor model: Comments on current theory and practice in second language acquistion.* Boston: Heinle & Heinle.

Swan, J. (1987). Some difficulties in applying a holistic expository rating scale to student narrative writing. *Kiyo 22:* 33-49. Osaka, Japan: Baika Women's College.