The Comprehension Hypothesis and its Rivals

Stephen Krashen

The main rivals to the input or comprehension hypothesis are the skill-building and comprehensible output hypotheses. Studies claiming to show that the skill-building hypothesis is correct consist of studies of the impact of grammar instruction and error correction. These studies show only that instruction and correction have a modest impact on tests in which subjects are focused on form. Comprehensible output is rare, and there is no evidence supporting its efficacy.

INTRODUCTION

There are several competing hypotheses concerning language acquisition:
(1) The comprehension (or input) hypothesis claims that we acquire language by understanding messages. More precisely, comprehension is a necessary, but not a sufficient condition for language acquisition. Other conditions must be met: an open attitude, or low affective filter, and the presence in the input of aspects of language that the acquirer has not yet acquired but is developmentally ready to acquire ("i+1"). Research in a variety of areas supports this claim.
(2) The skill-building hypothesis claims that we must first learn the rules of language consciously and they become automatic through output practice and correction.   I will argue that current research claiming to show that this hypothesis is correct is consistent with the Monitor hypothesis, which claims that the use and application of conscious knowledge is severely limited.
(3) The comprehensible output (CO) hypothesis claims that we acquire language by attempting to use new structures and vocabulary in production; we acquire when we achieve communicative success. I will argue that this hypothesis is incorrect: CO is too infrequent to make any significant contribution to language competence, and no study shows that those who experience more CO make better progress than those who experience less.

THE INPUT HYPOTHESIS

Evidence for the Input Hypothesis remains very strong. In second and foreign language acquisition method comparisons, students in classes containing more comprehensible input consistently outperform those containing less: At the beginning level, students in Total Physical Response classes, and Natural Approach classes do better than students in traditional classes on communicative tests, and do at least as well, or better, on grammar tests (Krashen, 1982, 1994). At the intermediate level, students in sheltered classes (content courses made comprehensible for intermediate students) do as well or better than comparisons (and acquire subject matter at the same time) (Krashen, 1994).

Additional evidence for the correctness of the comprehension hypothesis comes from research on recreational reading, written comprehensible input that is clearly accompanied by a low affective filter. Studies in both second and foreign language confirm that those who read

more do better on a wide variety of tests, and those who participate in sustained silent reading programs do better than those in traditional classes (Krashen, 2002).

THE SKILL-BUILDING HYPOTHESIS

There are two main competitors to the Input/Comprehension Hypothesis. The first is an old foe, the Skill-Building hypothesis, and the second, the Comprehensible Output hypothesis, is a new foe, but one that has roots in a common view among amateurs, the idea that we acquire language when we "need" to communicate.

The Skill-Building hypothesis claims that we acquire language by first consciously learning the grammatical rules, and by practicing applying the rules when producing output. Error correction is an important part of skill-building: When we are corrected, we are supposed to change our provisional conscious rule.

For the public, the Skill-Building hypothesis is not a hypothesis but an axiom; it is considered to be obviously true. I review here two sets of studies that help us evaluate the Skill-Building hypothesis. In both cases, I argue that the results confirm the severe limits of grammar instruction, are fully consistent with the Monitor hypothesis, a hypothesis that accompanies the Input Hypothesis as part of a general theory of language acquisition, and thereby confirm that comprehensible input is the major cause of language development.

The Monitor hypothesis claims that consciously learned grammar rules are available only as Monitor or editor (Krashen, 1982). To use the Monitor, three conditions must be met:

1. The acquirer must know the rule. This is a formidable condition. The rules of language are many and some are very complex.
2. The acquirer must be thinking about correctness, or focused on form.
3. The acquirer must have time to retrieve and apply the rule. In ordinary conversation, there is little time for rule retrieval and application, but there is more time in some kinds of oral language activities, especially those used in classrooms and examinations.

Grammar Instruction

Grammar instruction consists of two components: (1) focus on form and (2) presentation of a rule or rules. It is possible to do (1) without (2), that is, focus students on form without presenting a rule. This is done in several studies reviewed here. It is not, however, possible to do (2) without (1): When we present a rule, we are also focusing students on form.

Component (2) can take one of two forms. In one version, the students are given the rule: this has been termed "deductive" grammar learning. In another version, the students are asked to try to figure out the rule on their own ("inductive" learning). The latter version is not "acquisition" because the goal is the discovery of an explicit rule.

In Krashen (2002) I argued that studies that attempted to demonstrate the efficacy of direct instruction in grammar showed only that grammar teaching has a peripheral effect. These studies have several characteristics in common:

1. Subjects were experienced adult language students, which means they were used to direct teaching of grammar, expected it, and had survived it.
2. Comparison groups had either no treatment at all, or received impoverished comprehensible input. In addition, it is clear in some of the studies that some of the

comparison students were in fact focused on form. These studies, thus, investigated only the impact of more direct instruction versus less, not "learning" versus "acquisition."

3. Only the short term effect of conscious learning was investigated.
4. In all studies, the three conditions for Monitor use were met on all measures used: Subjects had studied the rules that were being tested, were focused on form during the test, and had time to apply the rules they had studied.

The consistent result was instructed subjects did better, but in nearly all cases the effect was quite modest and some cases it was completely absent. The studies thus only show that more instruction means a bit more consciously learned competence, a conclusion that is consistent with the claims of the Monitor hypothesis (Krashen, 1982).

The Norris and Ortega Review

The crucial test of the efficacy of instruction is whether it has an effect for Monitor-free tests given a substantial amount of time after the treatment. No study has shown that consciously learned rules have an impact on Monitor-free tests over the long term. I review here studies that, according to Appendix A in Norris and Ortega (2000), evaluated the impact of direct instruction and used "free constructed response," measures that "required participants to produce language with relatively few constraints and with meaningful communication as the goal for L2 production (e.g. oral interviews, written compositions)." (p. 440). Of the 12 studies cited by Norris and Ortega that utilized such tests, I was able to locate 11.

Table 1: Studies of the effect of instruction on "response-free tests"

	rule	treatment	delay	other tests FF	time pressure	comparison group(s)	results
Mackey & Philp	no	recast	5 wks	no	info gap	interaction/none	d = 1.53 (see text)
Salaberry	yes	input FF	1 mtn	yes	written	expl + output/none	no diff; few responses
Doughty (b)	yes	enh text	none	yes	pict descr	enh, reg text	all tests combined
Jourdinais et al	no*	enh text	none	no	written test	unenh text	d = 2.2, due to omissions
Van Patten & Sanz	yes	input FF	none	yes	pict desr	none	ns, but d = .58
White et al (b)	yes		5 wks	yes	info gap	direct instr.	41% > 63%**
Day & Shapson (a,b)	yes		11 wks	yes	interview	reg. Instr.	d = .48
Lyster	yes		1 mnth	yes	pict descr	reg. Instr.	tu/vous only
Harley (a)	yes		0,3 mtn	yes	interview	reg. Instr.	10% (immediate) 3% (delayed)

**well-formed questions, compared to pre-test performance, not controls
* both groups had studied the rule before treatment began; no diff when omissions not included; enh = reminder
enh = enhanced
a: test similar to class activities
b: prompt
"free response" effect sizes are smaller than grammar test effect sizes (Norris and Ortega); a reasonable result.

Table 1 presents these studies. Norris and Ortega did not provide details of effect sizes for individual studies. I therefore calculated effect sizes from the original studies and included them in table 1. This was not possible in some cases, however, because several studies did not include sufficient data for effect size calculation. In these cases, when possible, I present other data showing the effect of the treatment, e.g. percentage gain.

Two studies were not included in table 1. Yang and Givon (1997) was a comparison between pidgin-like input in an artificial language and fully grammatical input. This study did not investigate the impact of instruction, as the methodology provided to both groups was identical. Nagata (1997) was a comparison between two different kinds of grammar teaching: Deductive and inductive. Both groups of students of Japanese as a foreign language were heavily instructed, focused on form, and given the target rule explicitly.

The Nature of the Test: Conditions for Monitor Use Met

There is good evidence that the conditions for the use of the conscious Monitor were met during the "free-response" tests. All subjects had had instruction (the treatment), and thus had some knowledge of the target rule, and the tests clearly focused subjects on form and allowed enough time for conscious knowledge to be accessed and applied. Even so, the impact of instruction was modest.

Focus on form . In some cases, students took other tests that clearly had a focus on form, which confirmed for subjects that the name of the game was accurate performance on certain structures. (Harley, 1989; Day and Shapson, 1991; Lyster, 1994; White, Spada, Lightbown and Ranta, 1991; VanPatten and Sanz, 1995; Salaberry, 1997).

In several cases, during the administration of the test subjects were actually prompted to use target forms when they did not initially do so (Day and Shapson, 1991, Doughty, 1991, White et. al., 1991) and in at least two studies (Harley, 1989: Day and Shapson, 1991) the test was similar in format to some class activities obviously designed to teach the target forms.

Absence of Time Pressure. In no case were subjects subjected to the time pressure that exists in regular conversation. On the oral tests, subjects were asked to describe a picture or sequence of pictures (Doughty, 1991; Nagata,1997; Lyster, 1994; VanPatten and Sanz, 1995) or engaged in "information gap" activities in which they asked questions of a native speaker (Mackey and Philp, 1998; White et. al., 1991), and in two studies, tests were written (Salaberry, 1997; Jourdinais, Ota, Stauffer, Boyson, and Doughty, 1995). In Harley (1989) and Day and Shapson (1991), the oral test was an interview, but no indication is given whether time pressure existed or not.

Immediate Testing. Posttests were not delayed long enough for subjects' their conscious knowledge to fade; the longest delay in any of the measures in studies using free response was three months.

Additional Concerns..Comparison groups in these studies did not have exposure to the target rules in any form; they were not in an acquisition-rich environment. Thus, these studies really compare the presence and absence of formal instruction, not acquisition versus learning.

In Mackey and Philp (1998), it was unclear whether there was a focus-on-form in the treatments given to any of the groups studied. One experimental group received "recasts", that is, each time the subject made an error (in question formation), it was repeated in correct form by a native speaker of English. Comparison students participated in interactions but received no recasts, and another comparison group had no treatment at all. It is difficult to claim that the recast treatment was form-focused. Getting an error "recast" may not promote as obvious a focus on form as more direct modes of instruction. No rule was given and no other form of direct instruction occurred. [1]

In Doughty (1991), all test results were combined into one score; scores for the "free response" test was lumped together with other tests that were highly form-focused. We thus have no idea if student responses on the free response test were the same as or were different from those on grammar tests. In Lyster (1994), subjects had instruction on the French conditional and the tu/vous distinction, but on the free response test Lyster only considered performance on the latter, a "structurally simple rule" (p. 280). On written tests, subjects failed to show any gains on the conditional.

Jourdenais, Ota, Stauffer, Boyson, and Doughty (1995) is clearly not a study of rule acquisition or rule learning, but is a study of performance, of how much and how accurately students use a rule they had previously learned in class (the imperfect and preterit in Spanish). On a test given four weeks before the study, subjects averaged 8.14 correct out of ten in the section requiring them to fill in blanks with either the preterit or imperfect, indicating that they had a considerable amount of conscious knowledge of this rule before the treatment.

The "free response" test was a writing task done immediately after the treatment, which consisted of reading a paragraph with target forms enhanced (underlined and printed in a different font; in addition preterit verbs shadowed and imperfect verbs bolded). In fact, the writing task was printed on the same sheet of paper as the sample text described just above; thus, "enhanced" students still had the previous experience in mind when they did the writing task; they were still focused on form.

Those in the "enhanced" group were more accurate on the writing task in terms of correct forms supplied in obligatory contexts, but this result was largely due to the low performance of two subjects in the non-enhanced group. Because of the two low scorers, the effect size was very high in favor of the enhanced group (d = 2.16, calculated from p-values). When input-only subjects attempted to use the preterit or imperfect, they got it right just as often as did enhanced subjects. In other words, the "enhanced" group's higher accuracy, was due to the fact that they

---

[2] Mackey and Philp defined progress as subjects' use of a more advanced form of English questions two or more times in <u>any</u> of three different administrations of the posttest. The "recast" group made better progress: seven out of nine recast subjects "advanced" but only one out of six of the interaction-only advanced. Mackey and Philp report that according to a chi square analysis, this is statistically significant and according to my calculations it results in a substantial effect size (d = 1.53). But the results are nevertheless marginal. With such a small sample, a more appropriate statistical test is the Fisher test. According to the Fisher, the difference is still significant (p = .04). But this result is very fragile. If one fewer recast subject had made progress, according to the Fisher test the difference would not have been significant (p = .12). Mackey and Philp also dismissed the performance of the control group that received no treatment, merely stating, in a footnote, that the comparison of controls and recast subject was not "of interest." Two controls increased, four did not. According to the Fisher test, the difference between recast and control subjects was not statistically significant at the .05 level (p = .08). I am not one to regard the .05 level as an absolute standard, but the small sample size and marginal results do not inspire confidence. Even if recasted subjects did improve more than the others, it is not clear why they improved. Mackey and Philp, in fact, suggest that the recasts functioned "as part of a data-base for the language learner" (p. 353). It could have served, in other words, as comprehensible input. (It should also be noted that only 26% of the recasts were followed by subjects' repeating or modifying their own output, suggesting that subjects tended not to regard the recasts as corrections.)

had fewer omissions. Each group was equally accurate when they tried to actually write the appropriate verb, confirming that enhancement simply reminded students to use a verb.

Both of Salaberry's experimental group received heavy focus on form. The rule (clitics in Spanish) was explained to both groups, but one had input-oriented grammar exercises, and the other output oriented. There was no difference in among the two experimental groups and a third group that received no instruction at all. The sample size was very small, with nine in the input group (only six took the follow-up test), 10 in the output group and seven in the control group. Subjects, however, produced very few object pronouns on the posttest.

Free-response Studies: Summary

Subjects who take "free response" tests are indeed focused on form, thanks to the treatment, other measures, and the nature of the tests themselves. It is thus no surprise that one sees some impact of instruction on these tests: the conditions for the use of the Monitor are met.

The Monitor hypothesis predicts that one will see more impact of instruction on tests when conditions for the use of the Monitor are met more fully. Thus, the free response tests should show less impact of instruction than performance on what are clearly discrete-point grammar tests in which there is more focus on form.

Norris and Ortega reported a smaller mean effect size ($d = .55$) for studies using "free response" measures than for studies using measures involving more focus on form (for studies using tests with "constrained responses", mean $d = 1.2$), which is consistent with the predictions of the Monitor hypothesis.

The Norris and Ortega review should not be interpreted as a triumph for direct instruction, but as fully consistent with the Monitor hypothesis.

Error Correction

As noted earlier, correction is a form of inductive grammar teaching. Correction encourages the language learner (not acquirer) to adjust the conscious rule he or she is using, or, if the rule is completely unknown, to try to discover a rule.

If this assumption is correct, there are severe constraints on learning rules from correction.
1. The learner must be thinking about rules, or focused on form immediately after the correction.
2. The learner must have enough time to think about grammar rules.
3. The rule must be learnable. Not all rules can be consciously learned.

Clearly, conditions (1) and (2) are met most completely when students do traditional grammar exercises in which it is obvious that the goal is the mastery of a rule. One would expect the above conditions to be hardest to satisfy during genuine conversation, but no study, to my knowledge, has studied the effect of correction during ordinary conversation.

Even if error correction has been successful, if the learner did indeed improve his or her conscious knowledge of a rule or discover a new rule, it will be hard to demonstrate this knowledge. There are also severe conditions on the use of conscious rules. In addition to knowing the rule,
(1) The learner must be thinking about rules, or focused on form while using the language.
(2) The learner must have time to retrieve and apply the rules.

These are, of course, the conditions for Monitor use. Once again, these conditions will be met most fully on a grammar test.

The effect of conscious learning fades with time. Just how long it takes for learned rules to be forgotten appears to depend on the rule and the amount of training, but studies have shown a weakening of the impact of learning after three months (Krashen, 2002).

It is widely accepted that language learners differ a great deal in their ability to learn rules and their interest in learning rules. Some are extremely sophisticated learners, with a keen interest in the structure of language and a belief that conscious learning is important. At the other extreme are second language students with little or no formal education. We would expect correction to have a greater impact on those with more interest in, and background in formal grammar learning.

Another variable is how focused the correction is. In some studies, all errors are corrected, in others, correction is restricted to a few target items. In the former case, all errors are considered in testing the impact of correction, in the latter only those focused on are tested. One would of course expect the impact of correction to be greater when correction is focused.

To summarize, for correction to be successful, conditions on the act of correction must be satisfied (focus on form, time, a learnable rule), and conditions on the use of conscious rules must be satisfied (focus on form, time). In addition, it is hypothesized that correction will tend to be more effective when learners are experienced learners, when the focus of the correction is on a small set of rules, and when the posttest is given soon after the treatment.

This review is limited to studies that allow the calculation of effect sizes. All studies are described in terms of the degree to which the constraints on correction are present, the degree to which the constraints on the use of conscious rules are present. When available, we also consider the effect of delaying the posttest.

In addition, if we are studying the impact of correction, we must only include studies in which groups differ only in that feature. Treatments must be identical, or very similar, the only difference being that one group is corrected and the other is not.

The prediction is that the impact of correction is greater when: Form-focus and time are present when the correction is done, correction is focused, form-focus and time are present when the impact of correction is tested, and the test is given soon after the treatment.

An ideal analysis would present from studies in which the presence or absence of correction was the only difference between experimental and control groups. A review of the literature produced only seven such studies that allowed the computation of effect sizes: (Carroll, Roberge and Swain, 1992; Carroll and Swain,1993; Tomasello and Herron (1998, 1989), DeKeyser (1993), Fazio (2001) and Mason (2002), which included nine comparisons. To expand the set of studies, studies in which correction was combined with supplementary grammar instruction were included. These studies should show a larger effect for correction than those that only compare the presence and absence of correction. Subjects are not only corrected but are also supplied with a rule, while comparison students have neither of these treatments. Table 2 presents both "pure" comparisons and studies in which correction was combined with supplementary grammar instruction and practice. Only the immediate posttest results are included in this table.

Table 2: Impact of Correction (expanded set)

Carroll et al (1992)	FF + time	focused	FF + time	0.56
Carroll & Swain (1993)	FF + time	focused	FF+ time	1.3
Tomasello & Herron (1988)	FF + time	focused	FF+ time	1.55
Tomasello & Herron (1989)	FF + time	focused	FF + time	1.02
DeKeyser (1993)	all activities	global	FF + time	0.64
Ramirez & Stromquist (1979)	FF + time	focused	FF + time	0.95
Cardelle & Corno (1981)	FF+ time	focused	FF + time	1.16
Jafapur & Yamini (1993)	dictation	global	FF + time	0.17
Ramirez & Stromquist (1979)	FF + time	focused	FF + time	0.95
Cardelle & Corno (1981)	FF+ time	focused	FF + time	1.16
Kepner (1991)	composition	global	composition	0.39
Sheppard (1992)	composition	global	composition	-0.68
Jafapur & Yamini (1993)	dictation	global	reading comp	-0.15
Polio, Fleck & Leder (1998)	composition	global	composition	0.13
Fazio (2001)	composition	global	composition	- 0.39
Mason (2002)	composition	global	composition	-0.25
Mason (2002)	composition	global	reading comp	0.08
Mason (2002)	composition	global	cloze	0.25

Examining only the variable of "test," for studies in which corrected students took grammar tests, correction produced consistently positive effect sizes, with a mean effect size of .919 (sd = .444). Those not using grammar tests had a mean effect size close to zero (mean = -.070, sd = ..356). These means were significantly different (t = 4.985, df = 14, p = .0002). There was nearly no overlap at all between the two groups, and none at all if we eliminate Jafapur and Yamini (1993)'s condition in which a grammar test was used; students in this study were corrected on dictations, and were not as heavily focused on form as students in other studies with heavily form-focused tests. These results are remarkably similar to the results reported for the restricted set of studies of "pure" correction. (This analysis includes all three conditions from Mason, 2002; including only one or taking the mean of the three produces very similar results.)

As noted, this analysis takes only the immediate posttest into consideration. Most of the studies that used grammar tests as posttests also employed focused correction and also focused students on form during correction. Also, all of the studies that did not use grammar posttest employed global correction, and none focused students on form during correction. It is thus very possible that these other factors are important. Studies systematically varying these conditions would be useful in determining if these other factors are relevant. Nevertheless, it is striking how well the posttest alone predicts the impact of correction.

The small amount of data available on the effect of delaying the posttest confirms that the conscious knowledge gained through correction fades in a short time. As noted in table 3 we see a decline in the effect size after only one to three weeks when the final tests is a grammar test.

Table 3: Results of delayed posttests

study	measure	effect size	delayed ES	length of delay
Jafapur & Yamini	listening comp	-0.08	0.07	one semester
Carroll, Swain & Roberge	FF+time	0.56	0.31	one week

| Carroll & Swain | FF+time | 1.3 | 0.76 | one week |
| Tomesello & Herron | FF +time | 1.02 | 0.78 | 17 days |

As noted, this is a preliminary analysis, based only on the posttest. Nevertheless, the conclusions are in clear agreement with those of Truscott (1996), who concluded that explicit correction of grammatical errors had no effect on writing in a second language, and that studies in which correction appears to have an effect utilized tasks and tests in which students were heavily focused on form (pp. 340-41). It appears to take a grammar test to show the effects of correction. We do not yet know if the zero impact of correction on other tests is due to a lack of learning during correction or an inability to display this knowledge.

The review of research on grammar teaching and correction both confirm the limits of conscious learning. Correction, however, even when accompanied by direct instruction, appears to fail completely on anything except a grammar test. As noted in the previous section, grammar instruction has a modest effect even on "free response" tests, tests with a lighter focus on form than grammar tests. An obvious reason for this is the far greater intensity of treatment in studies categorized as grammar instruction studies.

Comprehensible Output?

The comprehensible output (CO) hypothesis states that we acquire language when we attempt to transmit a message but fail and have to try again. Eventually, we arrive at the correct form of our utterance, our conversational partner finally understands, and we acquire the new form we have produced. The originator of the comprehensible output hypothesis, Merrill Swain (Swain, 1985), does not claim that CO is responsible for all or even most of our language competence. Rather, the claim is that "sometimes, under some conditions, output facilitates second language learning in ways that are different form, or enhance, those of input" (Swain and Lapkin, 1995, p. 371). Even this weak claim, however, is hard to support.

A problem all output hypotheses have is that output is surprisingly rare (Krashen, 1994). In Krashen (2002), I concluded that in the case of CO, the problem is especially severe. When we examine studies in which nonnative speakers conversed with native speakers, instances of comprehensible output (in which native speakers indicate lack of comprehension and nonnative speakers respond with a more grammatically accurate utterance) are very infrequent, ranging from about one per hour to one every five minutes. The same appears to be true in classrooms.

The only published study claiming to show greater improvement for subjects receiving comprehensible output (Nobuyoshi and Ellis, 1993) involved only three subjects in the experimental condition. One of the three subjects showed no gain at all and in another the gain was insignificant. The claim that third subject gained is based on an analysis of only 13 utterances in one session and nine utterances during a subsequent session. In addition, all subjects had studied the target rule before the study began and were clearly focused on form during the treatment. It is thus not clear that improvement was due to new acquisition or learning; comprehensible output may have simply served to remind the subjects to use the Monitor.

Comprehensible output requires that acquirers be "pushed to use alternative means to get across .. the message ... precisely, coherently, and appropriately" (Swain, 1985, pp. 248-249). There is evidence that "pushing" students to speak is unpleasant for them. When asked what aspects of foreign language classes are the most anxiety-provoking, students put "talking" at the top of the list (Krashen, 2002). Laughrin-Sacco (1992) reported that for students in beginning

French classes, "for nearly every student ... speaking was the highest anxiety-causing activity" (p. 314). Ten "anxious" foreign language students interviewed by Price (1991) stated that a source of stress "was the frustration of not being able to communicate effectively" (p. 105). These results suggest that it is "pushed output," having to utilize structures they have not yet acquired, under demanding conditions, that students find uncomfortable. Methods based on comprehensible output put students in this situation constantly.

Given the consistent evidence for comprehensible input (Krashen, 1994) and failure of other means of developing language competence, providing more comprehensible input seems to be a more reasonable strategy than increasing output.

REFERENCES

Cardelle, M. and Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, 15(3), 251-261.

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

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

Day, R. and S. Shapson. (1991). Integrating formal approaches to language teaching in French immersion: An experimental study. *Language Learning,* 41, 25-58.

DeKeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal,* 77(4), 501-513.

Fazio, L. (2001). The effect of corrections and commentaries of the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing,* 10 (4), 235-249.

Harley, B. (1989). Functional grammar in French immersion: A classroom experiment. *Applied Linguistics.* 10, 331-359.

Jafarpur, A. and Yamimi, M. (1993). Does practice with dictation improve language skills? *System,* 21(3), 359-369.

Jourdinais, R., Ota, M., Stauffer, S., Boyson, B., and Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol. In R. Schmidt (Ed.) *Attention and awareness in foreign language learning* (pp. 183-216). Honolulu: University of Hawai'I, Second Language Teaching and Curriculum Center.

Krashen, S. (1982), *Principles and practice in second language acquisition*. Englewood Cliffs, NJ: Prentice Hall.

Krashen, S. (1994). The input hypothesis and its rivals. In Ellis, N. (Ed.) *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Krashen, S. (2002). Explorations in language acquisition and use: The Taipei lectures. Taipei: Crane Publishing Company.

Loughrin-Sacco, S. (1992). More than meets the eye: An ethnography of an elementary French class. *Canadian Modern Language Review* 49, 80-101.

Lyster R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' socio-linguistic competence. *Applied Linguistics,* 15, 263-287.

Mackey, A. and Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal,* 82, 338-356.

Mason, B. (2002). Sufficiency of extensive reading on the development of grammatical accuracy. Ph.D. Dissertation, Temple University, Osaka, Japan.

Polio, C., Fleck, C. and Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing,* 7(1), 43-68.

Nagata, N. (1997). An experimental comparison of deductive and inductive feedback generated by a simple parser. *System,* 25, 515-534.

Nobuyoshi, J. and Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal* 47, 203-210.

Norris, J. and Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-428.

Price, M. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In Horwitz,E. and Young, D. (Eds.), *Language Anxiety* (pp.101-108). Englewood Cliffs, NJ: Prentice Hall.

Ramirez, A. and Stromquist, N. (1979). ESL methodology and student language learning in bilingual elementary schools. *TESOL Quarterly,* 13, 145-158.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills: Sage.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal,* 23, 103-110.

Salaberrry, M. (1997). The role of input and output practice in second language acquisition. *Canadian Modern Language Review,* 53, 422-451.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass, S. and Madden, C. (Eds.), *Input in second language acquisition* (pp. 235-256). New York: Newbury House.

Swain, M. and Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics,* 16, 371-391.

Tomasello, M., and Herron, C. (1988). Down the Garden Path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics,* 9(3), 237-246.

Tomasello, M. and Herron, C. (1989). Feedback for language transfer errors: The Garden Path technique. *Studies in Second Language Acquisition,* 11, 385-395.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning,* 46(2), 327-369.

VanPatten, B. and Sanz, C (1995). From input to output: Processing instruction and communicative tasks. In F. Eckman, D. Highland, P. Lee, J.Mileham, and R. Weber (Eds.), *SLA theory and pedagogy* (pp. 169-185). Hillsdale, NJ: Lawrence Erlbaum.

White, L., Spada, N., Lightbown, P., and Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics,* 12, 416-432.

Wolf, F. (1986). *Meta-analysis: quantitative methods for research synthesis.* Beverly Hills: Sage.

Yang, L. and Givon, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition. *Studies in Second Language Learning,* 19, 173-194.