Evidence in Favor of a Broad Framework for Pronunciation Instruction

Tracey M. Derwing
University of Alberta

Murray J. Munro
Simon Fraser University

Grace Wiebe
University of Alberta

We had native English-speaking (NS) listeners evaluate the effects of 3 types of instruction (segmental accuracy; general speaking habits and prosodic factors; and no specific pronunciation instruction) on the speech of 3 groups of English as a second language (ESL) learners. We

recorded their sentences and extemporaneously produced narratives at the beginning and end of a 12-week course of instruction. In a blind rating task, 48 native English listeners judged randomized sentences for accentedness and comprehensibility. Six experienced ESL teachers evaluated narratives for accent, comprehensibility, and fluency. Although both groups instructed in pronunciation showed significant improvement in comprehensibility and accentedness on the sentences, only the global group showed improvement in comprehensibility and fluency in the narratives. We argue that the focus of instruction and the attentional demands on speakers and listeners account for these findings.

Although interest in instruction in pronunciation for ESL learners has increased dramatically in the last decade (Morley, 1994), very little evidence exists as to whether one focus of pronunciation instruction is superior to another or even whether any form of instruction is beneficial at all (Pennington & Richards, 1986). Teachers have had to rely on their own intuitions or those of materials developers to decide on the emphasis a pronunciation course should take. Numerous reports advocate one approach or method over another; for example, Brown (1995) made a convincing argument for the case that training with minimal pairs is far less useful than training in suprasegmentals, but he had no data to that effect. Several other accounts have championed a particular view (e.g., Acton, 1984; Firth, 1992; Griffen, 1991), but each had little or no supporting empirical evidence.

Some studies have looked at the short-term effectiveness of very limited training on various aspects of oral production. De Bot (1983), for example, found visual feedback effective as a means of improving intonation in Dutch EFL learners. However, he provided only limited training (45 or 90 minutes), and assessed only speech produced immediately after training. De Bot and Mailfert (1982), Gilbert (1980), and Neufeld and Schneidermann (1986) all reported studies with similar limitations.

In contrast with the studies cited above, Perlmutter (1989) found improved intelligibility in ESL learners during six months of language instruction with particular emphasis on pronunciation. However, because the students were newcomers to the United States, who were likely to show significant improvement regardless of instruction (Flege, 1988), one cannot determine how much, if any, of their improvement was directly attributable to the pronunciation-specific instruction.

Derwing, Munro, and Wiebe (1997) showed that long-term ESL individuals' pronunciation improved significantly in a 12-week program emphasizing global production skills. The improvement was evident to untrained raters, who performed blind judgments of short speech samples, recorded before and after instruction. Some of the ESL learners also demonstrated significant improvement in intelligibility, determined by native listeners' transcriptions of the ESL students' speech samples. Although the outcomes of this study were promising, the instructors restricted their lessons to the type of global strategies (prosodic features, general speaking habits) advocated by Firth (1992) and Gilbert (1993). Consequently, we had no way of knowing whether such an emphasis would be the most effective in leading to overall improvement in ESL pronunciation. Because many materials and programs are primarily segment-based—that is, they focus on the perception and production of individual phones—it is important to compare the types of content used in different pronunciation courses.

Macdonald, Yule, and Powers (1994) compared the productions of ESL students under 3 types of instruction—teacher correction, self-study in a language lab, and interactive modification—with those of an ESL control group who received no treatment. They assessed the students' productions before the experimental treatment, immediately after, and again 2 days later. Although the results were most positive in the self-study condition, the authors acknowledged that their findings were far from conclusive. As they pointed out, there was no indication whether

improvement could be sustained over time; the groups were very small; and the self-study group had more time than students in the other groups.

To extend the research cited above we undertook a study comparing the implementation of 3 perspectives on pronunciation teaching over a much longer period of time. We collected "before and after" speech samples from 3 groups of ESL learners: one taught with a segmental focus, a second taught with a global focus, and a third that received no specific pronunciation instruction. To make our investigation comprehensive, we asked the learners to read sentence-length utterances as well as to produce extemporaneous narratives to approximate a more "real-life" speaking condition (see Derwing & Munro, 1997; Munro & Derwing, 1995a). English NS listeners evaluated the speech samples by rating them for accentedness, comprehensibility and fluency in blind tasks.

Several rating studies have explored in considerable detail the dimensions of accentedness, comprehensibility and intelligibility (Derwing & Munro, 1997; Munro & Derwing, 1994, 1995a, 1995b). Accentedness refers to the extent to which a listener judges second language (L2) speech to differ from NS norms. Research has repeatedly shown that even heavily accented speech can be highly comprehensible. Comprehensibility in this context is the listener's judgment of how difficult it is to understand an L2 speech production. Researchers have typically obtained judgments of both these dimensions via rating scales. Comprehensibility is a subjective assessment of ease or difficulty of comprehension as opposed to a measure of actual intelligibility. The latter refers to how much of an utterance the listener processes successfully; it may be quantified through comprehension questions or an orthographic transcription task (e.g., *dictée*). Finally, one can also assess fluency through scalar ratings. In our present study, fluency refers to overall tempo and flow as opposed to a measure of general proficiency (cf. Lennon, 1990).

Prabhu (1990, 1992), among others, identified some of the problems associated with an experimental comparison of methods.

One of his primary concerns was the degree of investment the teacher has in the method to be evaluated, most especially the "teacher's sense of plausibility" (1990 p. 172): in other words, the teacher's internal notion of how learning takes place. Our present research addressed this issue by seeking a cooperative relationship with teachers who understood the purpose of the study and who wanted to participate.

A second concern relates to the construct, "method." In the early 1960s, a number of studies explored the hope that one method of L2 teaching would prove superior to all others. In general, such studies entailed examining learner outcomes in classes in which the same or similar content was taught through a variety of methods. As several researchers have pointed out (e.g., Allwright & Bailey, 1991; Long, 1980), the search for a "best" method was fraught with problems because of complexities within the classroom.

In the present study, although it may appear that we compared 2 methods of pronunciation instruction, we in fact compared 3 conceptions of pronunciation pedagogy. In one, we focused on instruction at the level of the word or smaller units; in the second, we emphasized a broader interpretation, extending to the level of interactive discourse. We compared these two conceptions of pronunciation (narrow vs. broad) with a third, the laissez-faire approach (i.e., no pronunciation-specific instruction). In traditional terms, students in this third category of instruction served as a control group. Rather than comparing methods, then, we compared a difference in the scope of content in pronunciation classes; indeed, the teachers in this study employed some of the same techniques, emphasizing both receptive and productive skills. For instance, both conceptions entailed repetition. However, the segment-based approach involved the elicitation of individual sounds and syllables; the global approach focused on larger units incorporating stress, intonation, and rhythm.

Experiment I: Sentence Data

Method

Speakers. Forty-eight adult students at the intermediate proficiency level in a full-time ESL program participated in the study. They ranged in age from 18 to 44 years, with a mean age of 31.7 years. Their mean length of time in English-speaking Canada was 3.4 years; the range was 7 months to 15 years. We assigned participants to one of 3 groups of 16, roughly balanced for L1, gender, age on arrival, and length of time in Canada. Table 1 provides details on each group. The students clearly invested in the project; they showed evident interest when they met with us at the beginning and end of the course. After discussing the experiment with us, they appeared to understand the project's purpose. All of them volunteered to participate in the data collection, even though they had the option of going to the library during this time period. They asked many questions, and most signed a sheet requesting a summary of the results.

Teachers. Prior to the beginning of the courses, Derwing met with a college administrator to identify highly qualified teachers

Table 1

Speaker Group Information

	NSP group	Global group	Segmental group
L1:			
E. European	8	10	9
E. Asian	4	2	1
Spanish	2	3	3
Other	2	1	3
Gender	6 male, 10 female	4 male, 12 female	6 male, 10 female
Age of Arrival	$M = 29.05$	$M = 26.98$	$M = 27.45$
Time in Canada	$M = 3.19$	$M = 2.71$	$M = 3.86$

who might be willing to participate in the study. Once the administrator nominated 3 interested individuals, Derwing met with them and explained the nature of the research. She suggested that both the segmental and global approaches were equally likely to yield positive results. The teachers self-selected for the approach they preferred to teach; the control teacher had not studied linguistics and did not feel comfortable teaching either the global or segmental approach. The other two teachers had both taken linguistics courses in addition to their TESL training and were confident in their abilities to teach the method they chose. We then negotiated to determine the course content and procedures. Derwing met the instructors again at the half-way point to ensure that all was going well. They expressed a willingness to continue with the lessons and reported that their students appeared to be enjoying the course.

Listener-Raters. The 48 listeners in the sentence task were Canadian English NSs recruited from education classes at the University of Alberta. All had grown up in Canada west of Quebec, and none had learned to speak a second language fluently. All reported having normal hearing. A majority of participants reported having little or no regular contact with non-native speakers (NNSs) of English.

Instruction. The 3 conditions were that one group received no specific pronunciation instruction (NSP), one received segmental instruction, and one global instruction. The NSP group attended skills-based (reading, writing, listening, speaking, grammar) ESL classes for 20 hours per week. The segmental and global groups also attended ESL classes for 20 hours per week, but their regular program included a daily pronunciation component. The global group received approximately 20 minutes per day of instruction in which the teacher focused on features such as speaking rate, intonation, rhythm, projection, word stress, and sentence stress. This teacher extensively used both materials (such as *Jazz Chants*, Graham, 1978; *Sounds Great*, Beisbier, 1995) and group presentations. The instructor used commercial materials in novel ways. For instance, she would have the students

count the number of syllables and number of stresses in each line of a Jazz Chant. The students would tap out the beats and use nonsense syllables to focus on rhythm (as suggested by Celce-Murcia, Brinton, & Goodwin, 1996). The teacher made no attempt to focus on individual consonant and vowel sounds. The segmental group used language-lab materials in conjunction with teacher-led exercises, again for approximately 20 minutes per day, that were designed to improve their productions of individual sounds (Corbett, 1992). Activities included identification and discrimination tasks as well as repetition tasks featuring individual sound contrasts.

Speech Samples. We collected speech samples from all ESL participants near the beginning of their course (Time 1) and again 11 weeks later (Time 2). They recorded their utterances on a Sony Console LLC-9000 system in a college language laboratory. The first item was a list of simple statements. We chose these items because we had previously found them suitable for sentence-level accent and comprehensibility judgments (see Munro & Derwing, 1995b). All the sentences consisted of a single clause and all contained high-frequency lexical items (e.g., "Many people drink coffee for breakfast," or "You can start a fire with a match"). In preparation for the recording, first the students read the sentences silently. Then, one of us read the complete set of statements aloud while the students followed silently. Once we had answered any questions about unfamiliar words, the students read the sentences aloud onto tape. Immediately afterward, they recorded the speech sample for the second experiment. At no point between Time 1 and Time 2 did the teachers review or use as teaching material the sentences or picture stimuli.

Control Recordings. We made additional recordings of 4 Canadian English NSs (2 male, 2 female). They served as a check on individual listeners' use of the rating scales in the listening tasks. We expected that all raters would consistently assign very good scores to the NSs. Failure to do so might indicate a misunderstanding of the instructions.

Stimulus Preparation. We used the speech materials described above to produce stimuli for the speech evaluation tasks. So that the stimuli could be conveniently randomized and presented to the raters, we digitally recorded them and saved them as audio files on a Macintosh computer at 22kHz with 16-bit (i.e., CD-quality) resolution. We then rerecorded the stimuli on tapes for random presentation during the 2 listening sessions. For the sentence rating tasks, we used two statements from each speaker, one from Time 1 (the beginning sample) and the same statement from Time 2 (the final sample). Eight statements produced by the NSs functioned as a check.

Procedure. Participants completed the experiment in several group listening sessions held over a period of 2 weeks. To reduce the likelihood of fatigue, we required each listener to complete the task on 2 separate days. The listeners heard the stimuli on a Sanyo MCD-Z31 stereo tape deck in a classroom and rated them using 9-point scales in the same manner as in previous, related studies (e.g., Derwing & Munro, 1997). After playing each stimulus once, we provided a pause, during which the raters decided how difficult the utterance was to understand and circled a comprehensibility rating from "1" (very easy to understand) to "9" (impossible to understand). One of us manually controlled the pause time to ensure that all listeners were comfortably in step. The average pause time was about 4 seconds. Immediately afterward, the listeners evaluated how accented the same utterance was and circled an accent rating from "1" (no accent) to "9" (very strong accent). We advised the listeners to use the entire scale. We presented them with three warm-up items at the beginning of the task. Each participant received an honorarium of $20.00 at the end of the second session.

Results

We followed commonly accepted practice in excluding data from a few listeners from the analysis. One of these listeners

assigned several ratings of 3 and 4 to the NS talkers. None of the other listeners displayed such a tendency. In fact, the mean values for the NSs were 1.06 for comprehensibility and 1.05 for accent. We excluded 4 other listeners because they failed to use the full scales as instructed. We assessed inter-rater reliability by taking data from a random selection of 15 of the remaining 43 listeners. We transformed inter-rater correlations into Z-scores and calculated the mean, following the procedure described by Hatch and Lazaraton (1991). The analysis yielded Pearson coefficients (r) of .71 and .70 (indicating a moderate level of inter-rater agreement) for the comprehensibility ratings and accent ratings, respectively.

By pooling over listeners, we computed mean comprehensibility scores for each speaker. An examination of these values revealed that some of the speakers received mean scores lower than 3.0. We decided to omit these people's data from the analysis for two reasons: first, to ensure that the 3 groups were comparable at Time 1, by setting a common lower limit for all 3 groups, and second, to reduce the likelihood of floor effects (lower scores indicated better performance). In particular, it seemed improbable that a participant who obtained nearly NS ratings (ratings of 1 or 2) from the listeners at Time 1 would show much evidence of improvement as a result of any type of instruction. In total, we excluded 3 speakers from the NSP group, 5 from the global group, and 4 from the segmental group.

We submitted the comprehensibility data from each listener to a two-way mixed design ANOVA with Time (1 or 2) and teaching focus (segmental, global, or NSP) as factors. A p level of .05 applies in all analyses reported below unless otherwise indicated. The effects of focus, $F(2, 84) = 8.1$, and time, $F(1, 42) = 48.8$, were significant, as was the interaction of the two factors, $F(2, 84) = 23.0$. Tests of simple main effects indicated no differences among groups at Time 1. However, both the global and segmental groups improved significantly from Time 1 to Time 2, though the NSP group did not. Tukey (Honestly Significant Difference) tests indicated that at Time 2, there was no difference between the global

and segmental groups, though both had significantly more comprehensible productions than the NSP group.

We submitted the listeners' accent judgment data to an ANOVA parallel to the one used for the comprehensibility data. Once again, the effects of focus, $F(2, 84) = 3.3$, and time, $F(1, 42) = 62.8$, were significant, as was the two-way interaction, $F(2, 84) = 7.7$. Tests of simple main effects indicated no significant difference among the groups at Time 1 and significant improvement in accentedness for all 3 groups from Time 1 to Time 2. However, at Time 2, the segmental group was significantly better than both the NSP and the global group. Thus, the segmental group showed the greatest overall improvement in accent in the sentence production task. Thomas Scovel (personal communication, March 27, 1997) noted the segmental group's apparently greater length of residence in Canada (see Table 1). Because this difference might explain this group's superior performance on the sentence task, we carried out a t test to determine whether, indeed, there was a significant difference in mean length of residence in Canada between the global and segmental groups (using data from only those participants included in the other analyses). The results proved nonsignificant, $t(21) = 1.28, p > .05$. These results and those of Experiment 2 are summarized in Table 2.

Table 2

Effects of instruction on performance in the three groups

	NSP	Segmental	Global
Sentences			
Comprehensibility	no change	improved	improved
Accentedness	improved	improved[a]	improved
Narratives			
Comprehensibility	no change	no change	improved
Accentedness	no change	no change	no change
Fluency	no change	no change	improved

[a]significantly greater improvement than the global and NSP groups

Experiment II: Narrative Data

Method

Participants and Recordings. The speakers were the same 48 ESL students as in Experiment 1. We have already described the recording procedures (see Sample Collection, Experiment 1).

Speech Sample. The speech sample for this experiment, recorded immediately after the sample for Experiment 1, consisted of an extemporaneous narrative description of a standard picture story we had used previously (see Munro & Derwing, 1994, 1995a; Derwing & Munro, 1997). For the narrative evaluation tasks, we selected two 45-second excerpts from the beginning of each student's extemporaneous description, one from Time 1 and one from Time 2.

Listeners. In this experiment we chose to recruit a smaller group of listeners, chiefly because of the task's high demands. We required the judges to listen to 45-second segments from each speaker in 2 listening sessions lasting a total of 5.5 hours. The judges were 6 experienced women ESL teachers, all of whom had familiarity with learners from a broad range of L1 backgrounds at all proficiency levels. All had normal hearing according to self-report. Derwing was one of the judges. Because we blinded the listeners to the time and condition of the speech samples, it is highly improbable that unconscious bias could have affected her ratings.

Procedure. We held the listening sessions over 2 days, with regular breaks to alleviate fatigue. The judges assigned comprehensibility and accent ratings in the same manner as the untrained listeners had in Experiment 1's sentence-rating task. In addition, they rated each speech sample for fluency on a scale ranging from "1" (NS-like fluency) to "9" (extremely disfluent).

Results

We assessed inter-rater reliability for the 6 judges using the procedure described earlier. We computed a Pearson coefficient (r)

of .72 for the comprehensibility scores. For the accent scores the value was .69, and for the fluency scores it reached .74. Even though the judges in this task were all experienced teachers, these values did not differ appreciably from those given by the untrained listeners in Experiment 1.

By pooling scores across judges, we obtained mean comprehensibility scores for each speaker. This time, 2 participants in the global group and 4 participants in the segmental group received ratings lower than 3.0. Once again, to ensure that all groups were comparable at Time 1 and to reduce the likelihood of floor effects, we omitted these speakers from the analysis.

We submitted the comprehensibility judgment data to a 2-way ANOVA. Although the overall effect of teaching focus, $F(2, 39) = .75$, failed to reach significance (our significance level was again $p < .05$), the effect of time was significant, $F(1, 39) = 5.1$. However, tests of simple main effects indicated that, although there was no difference among groups at Time 1, only the global group exhibited significant improvement in comprehensibility at Time 2.

A parallel analysis of the accent judgments revealed nonsignificant effects for both focus, $F(2, 39) = .8$, and time, $F(1, 39) = 2.5$. There was no evidence, then, that the training had any effect on the accent scores.

Although all groups tended to improve in fluency, as might be expected in any type of ESL program, tests of simple main effects indicated that the only group showing significant improvement in fluency was the global group. Once again, there was no difference among groups at Time 1. For a summary of the results of both this and Experiment 1, see Table 2.

General Discussion

These 2 experiments indicated that three aspects of oral production—comprehensibility, accent, and fluency—showed improvement as a result of instruction. Moreover, both the unit of measurement (read sentences vs. extemporaneous narratives)

and the focus of pronunciation instruction used (segmental vs. global) had an impact on the findings.

In Experiment I, which employed sentence judgments, the two pronunciation-specific groups evidenced similar improvements in comprehensibility, in contrast to the no instruction group, which demonstrated no significant improvement. Although accent scores for all three groups improved, the listeners judged the segmental group to be significantly less accented at Time 2 than the other groups. At first blush, this might suggest that a segmental focus would be marginally preferable to a global focus. However, the results of Experiment 2 indicated otherwise. In the extemporaneous narratives, the only clear evidence of improvement in comprehensibility and fluency was in the global group. None of the groups showed any noticeable improvement in accentedness.

Pronunciation instruction with only segmental content seemed to require that the student direct attention to the forms in question. When the nature of the linguistic task necessitated that attention be divided amongst lexical access, syntactic well-formedness, phonological accuracy, discourse organization, and so forth, speakers could not allocate enough resources to phonological concerns for there to be a noticeable transfer of segment-based skills. On the other hand, speakers who had had instruction emphasizing prosodic features such as rhythm, intonation, and stress could apparently transfer their learning to a spontaneous production.

Another way of interpreting the results is to consider the listeners' point of view. In the sentence rating task, listeners might have been better able to notice differences in accentedness because the possible number of errors was relatively small. Therefore, a segmental error was likely to be more salient in a single, grammatically correct, short sentence than in a 45-second-long sample of speech containing many competing errors and error types (grammatical, phonological, fluency, discourse, etc.). Two other studies support this account, both of which used short extemporaneous utterances as rating stimuli. Munro and Derwing (1995a) reported more significant correlations between error types

and accent scores in advanced learners than did Derwing and Munro (1997) in intermediate learners (i.e., learners who produced many more errors).

Moreover, the professional judges in this study reported some difficulty in coming to a decision on the accentedness scores. They attributed this to within-speaker variation across the 45-second narratives. No such difficulty occurred in the sentence rating task. Furthermore, in our numerous other studies using short utterances, NS listeners have never appeared to have difficulty assigning accent scores, even when allowed only a brief interval of 2 seconds to do so.

We elicited fluency ratings only in Experiment 2. Given that Experiment 1's sentence production task allowed for student preparation and required no extemporaneous generation of language, we anticipated that students' performance would vary only minimally in fluency. For this reason we chose not to collect fluency ratings from the (untrained) listeners. Although all 3 groups showed a trend toward increased fluency at Time 2 in the narrative task (which one would expect of learners in a general ESL program), only those students in the global group improved significantly. We can directly attribute this effect to the type of pronunciation instruction they received. As pointed out by the instructor of the segmental group in a debriefing after the final data collection, it was not surprising that a course focusing on individual sounds (as opposed to one dealing with larger units of speech) would fail to enhance students' fluency.

These findings do not necessarily speak in favor of a wholesale abandonment of a segmental focus in pronunciation teaching. Rather, we argue that attention to both global and segmental concerns benefits ESL students. In the case of a communication breakdown caused by a mispronunciation, a student who has received segmental training might be able to focus on the mispronounced form in a self-repetition. On the other hand, global instruction seems to provide the learner with skills that can be applied in extemporaneous speech production, despite the need to allocate attention to several speech components.

Although there have been many calls for studies of this sort (e.g., Morley, 1991; Pennington & Richards, 1986), ours is apparently the first controlled investigation of the effects of ESL pronunciation instruction over an extended period of time (a 10-week instructional interval vs. 2 days for Macdonald et al.'s study, 1994). Although this is a first step, many areas still require exploration, for example, the effect of instruction on learners of different ages.

Notwithstanding the many logistical difficulties in carrying out studies of this type, it behooves researchers to consider such factors as the efficacy of pedagogical materials and combinations of approaches, as well as the impact of instruction on learners of varying proficiency levels and L1s. Given the growing emphasis on pronunciation of late, we look forward to a clearer empirical identification of useful and effective approaches. Much current practice rests on intuitions and anecdotal evidence supplied by ESL practitioners, without even a modicum of empirical support. Although their intuitions may well be correct, it is necessary to find independent evidence through blind ratings by disinterested listeners. In the final analysis, ordinary listeners are the people to whom ESL students must be comprehensible.

<div align="right">Revised version accepted 27 March 1998</div>

References

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly*, *18*, 71–85.

Allwright, D., & Bailey, K. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers.* Cambridge: Cambridge University Press.

Beisbier, B. (1995). *Sounds great: Intermediate pronunciation and speaking* (Book 2). Boston: Heinle & Heinle.

Brown, A. (1995). Minimal pairs: Minimal importance? *ELT Journal*, *49*, 169–175.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.

Corbett, K. (1992). *Better English pronunciation for immigrants: Consonant and vowel corrections*. Calgary, AB: Coralex Publisher.

de Bot, K. (1983). Visual feedback of intonation: I. Effectiveness and induced practice behaviour. *Language and Speech*, *26*, 331–350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, *16*, 71–77.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19,* 1–16.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning, 8,* 185–203.

Firth, S. (1992). Pronunciation syllabus design: A question of focus. In P. Avery & S. Ehrlich (Eds.), *Teaching American English pronunciation* (pp. 173–183). Oxford: Oxford University Press.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America, 84,* 70–79.

Gilbert, J. B. (1980). Prosodic development: Some pilot studies. In R. C. Scarcella & S. D. Krashen (Eds.), *Research in second language acquisition* (pp. 110–117). Rowley, MA: Newbury House.

Gilbert, J. B. (1993). *Clear speech* (2nd ed.). Cambridge: Cambridge University Press.

Graham, C. (1978). *Jazz chants*: *Rhythms of American English for students of English as a second language*. New York: Oxford University Press.

Griffen, T. D. (1991). A non-segmental approach to the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings* (pp. 178–190). London: Routledge.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387–417.

Long, M. H. (1980). Inside the "black box": Methodological issues in classroom research on language learning. *Language Learning, 30,* 1–42.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve L2 pronunciation: The variable effects of different types of instruction. *Language Learning*, *44,* 75–100.

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly*, *25*, 481–520.

Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 66–91). Alexandria, VA: TESOL.

Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, *11*, 253–266.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech*, *38*, 289–306.

Neufeld, G., & Schneiderman, E. (1986). Prosodic development. In R. C. Scarcella & S. D. Krashen (Eds.), *Research in second language acquisition* (pp. 105–109). Rowley, MA: Newbury House.

Pennington, M. C., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly*, *20*, 207–26.

Perlmutter, M. (1989). Intelligibility rating of L2 speech pre- and postintervention. *Perceptual and Motor Skills*, *68*, 515–521.

Prabhu, N. S. (1990). There is no best method—why? *TESOL Quarterly*, *24*, 161–176.

Prabhu, N. S. (1992). The dynamics of the language lesson. *TESOL Quarterly,* *26*, 225–241.