A Frequency-Based Approach to L2 Phonological Learning: Teacher Input and Student Output in an Intensive ESL Context

PAVEL TROFIMOVICH, LAURA COLLINS, WALCIR CARDOSO, JOANNA WHITE, AND MARLISE HORST
Concordia University
Montreal, Canada

doi: 10.1002/tesq.9

■ Most second language (L2) researchers and teachers would agree that input, often defined as the language a learner hears or reads, plays an important role in L2 learning (see Piske & Young-Scholten, 2009). There is a great deal of research investigating which types of input are most beneficial for learning (e.g., Lightbown, 1985), how learners process and internalize input (e.g., Schmidt, 2001), and how input relates to language learning outcomes (e.g., Collins, Halter, Lightbown, & Spada, 1999). However, research investigating the role of input in learning L2 phonology, particularly in instructed settings, is conspicuous by its absence (although research in this vein with native speakers has proved fruitful, as shown by Bybee, 2002). In this article, we examine the relationship between teacher input and student output, focusing on the voiced interdental fricative /ð/ in English (as in *mother*).

Our data come from a large-scale longitudinal project investigating intensive English as a second language (ESL) programs in the French-speaking province of Quebec, Canada. Intensive ESL in Quebec is not content based. The regular curriculum (e.g., science, math) is completed in French in a condensed format, and the remainder of the school year (approximately 400 hours) is devoted to ESL. Typical ESL activities include skits, role-plays, and recitation of songs and poems, and often involve teachers reading aloud to students. Thus far, we have analyzed a 110,000-word corpus of teacher talk (Collins, Trofimovich, White, Cardoso, & Horst, 2009) and have examined lexical and morphosyntactic development in a sample of more than 200 learners (Collins & White, 2011). In this study, we extend our research to the domain of phonology by investigating the relationship between teacher input and student output with respect to /ð/, a difficult target for French learners of English (Trofimovich, Gatbonton, & Segalowitz, 2007).

Our analyses of the input-output relationship were informed by frequency-based approaches to language learning and use (Robinson & Ellis, 2008). Briefly, frequency-based approaches hold that language

users are sensitive to the frequency of lexical items in linguistic input and that language acquisition involves the learning of phonological, morphological, semantic, and other regularities from input. With respect to L2 phonology, the logic here is that certain aspects of speech (e.g., specific sounds, stress patterns, intonation contours) are easier to learn when they occur within and across a variety of recurrent familiar lexical items. The more frequently L2 learners experience a given phonological pattern in the input, especially across a range of lexical items, the more accurately they will perceive and produce this pattern.

Of course, frequency-based approaches are not the only theoretical perspectives that can be used to investigate the role of input in L2 learning. For example, input-output relationships could be studied within generative (Eubank & Gregg, 2002) or interactionist (Gass & Mackey, 2002) frameworks. These and other approaches ascribe a role for frequency in L2 learning but often vary in how important frequency is in explaining different L2 phenomena (for a review, see Ellis, 2002, and subsequent commentaries). In choosing to conceptualize this study within a frequency-based perspective, our overall objective was to determine the degree to which *frequency alone* could explain L2 phonological development, an aspect of L2 learning that has rarely been investigated from this perspective.

There is some evidence from L2 perception studies that frequency plays a role in L2 phonological learning. For example, Flege, Takagi, and Mann (1996) showed that Japanese speakers identified English /r/ and /l/ more accurately when these consonants occurred in frequent and presumably more familiar words (i.e., *room* and *lip* vs. *loom* and *rip*). Bradlow and Pisoni (1999) found that L2 learners recognized "easy" words more accurately than "difficult" ones, where easy words were high-frequency words with few lexical "neighbours" (similar-sounding words) and difficult ones were low-frequency words with many neighbours.

One shortcoming of these studies is that, in documenting the relationship between lexical frequency and language learning, the input learners actually received was not examined directly. Instead, researchers relied on frequency counts from language corpora, assuming that these counts reflected at least some of the properties of input. Another shortcoming is that these studies did not include frequency analyses of learner output, so it could not be determined whether the perceptual judgments were representative of how learners produced the target features, that is, how closely the frequency properties of the output matched those of the input. We addressed these limitations in longitudinal analyses of large corpora of teacher talk and learner language focusing on English /ð/ (which was never a target for explicit instruction). Our research objectives were (a) to compare the frequency profiles of teacher talk and learner output and (b) to examine how frequency

profiles of teacher talk relate to learner accuracy. Our overall pedagogical motivation was to understand whether properties of teachers' classroom talk can have important influences on learner language.

DATA AND ANALYSES

Teacher Input

Our original study of teacher input was based on a 110,000-word corpus that represented over 40 hours of aural input collected longitudinally in four Grade 6 intensive ESL classes (Collins et al., 2009). The video recordings were made at four 100-hour intervals of the 400-hour communicative program. For this study, we analyzed the recordings made on a single day at 100 and 300 hours of instruction because these recordings matched student output data (described below), also available at those times. The 57,818-word teacher corpus thus comprised the entire sample of three experienced teachers' speech, both spontaneous and scripted (e.g., in storybook readings).[1] Although two of the teachers were native French speakers, they were highly proficient in English.

The video recordings were first transcribed and then verified for accuracy by another transcriber. Next, a coder trained in phonetics listened to all instances of English /ð/ in the corpus and then classified them according to eight phonetic contexts. Based on Trofimovich et al. (2007), our assumption was that some contextual instances of /ð/ would be more frequent and therefore easier to learn than others. The contexts were (a) following vowels (*another*, *do that*), (b) following voiceless stops (*about this*), (c) following voiced stops (*read the*), (d) following voiceless fricatives/affricates (*pass them*), (e) following voiced fricatives/affricates (*means that*), (f) following nasals (*on those*), (g) following liquids (*before they*), and (h) sentence initially or phrase internally following a pause (… *the boy*, *girl's* … *that's*, where "..." designates a pause). After all /ð/ tokens were categorized, their frequency in each context was tallied separately for each teacher. The original classification was verified by another coder (agreement: 99%). Both coders identified all productions of English /ð/ in the teachers' speech as native-like.

Student Output

The student corpus consisted of 120 native French students' speech recorded at 100 and 300 hours of instruction (henceforth, Time 1 and

[1] One teacher taught two classes; the remaining two taught one class each.

Time 2). All students were 11- or 12-year-olds taught by the same three teachers. All students came from francophone communities outside Montreal and thus had little exposure to English outside of school. Before joining the intensive program, the students had received approximately an hour of ESL instruction per week during the previous 2 years of schooling (approximately 35 hours a year) and were therefore beginner-level learners of English.

Students met with a researcher in individual sessions that were recorded. Their task was to describe a picture sequence about two children going on a picnic (Muñoz, 2006). The researcher did not engage in conversation with students but, especially at Time 1, offered help with vocabulary or asked follow-up questions to elicit longer speech samples. All recordings were transcribed, then verified for accuracy. The 30,205-word student corpus ($M_{story\ length} = 126$ words) comprised only students' speech (excluding interviewer comments and asides).

The 240 picture stories (120 students $\cap$ 2 times) were analyzed independently by five coders trained in phonetics, such that every /ð/ token was evaluated by two different coders. The coders listened to each token and categorized it according to the same eight contexts described previously. The coders also rated each token for accuracy (1 = *sounds like a good English /ð/* or 0 = *does not sound like an English /ð/*). The coders were highly consistent in their classifications (Cronbach's $\alpha = .94–.99$) but showed less agreement in their accuracy ratings (Cronbach's $\alpha = .81–.96$). To minimize idiosyncratic effects of any individual coder, token counts and accuracy ratings were computed by averaging the two coders' responses.

RESULTS

Teacher Input

The goal of the input analyses was to document the properties of teacher input to students in order to explore possible relationships between teacher input and student output. To determine whether our teacher corpus was generally representative of English language use, we also compared teacher talk to 2,500 words of randomly selected texts from the British National Corpus (BNC).[2] Each teacher produced between 348 and 1,136 /ð/ tokens at each testing time, for a

[2] The BNC's speech corpus features written transcriptions of conversations, which makes it impossible to evaluate spoken instances of English /ð/. Instead, we chose to analyze a random selection of texts to get a sense of how /ð/ was distributed in a large corpus. Our 2,500-word selection followed the structure of the BNC (90% written texts, 10% transcribed conversations).

corpus-wide total of 4,778 tokens. When raw frequencies were adjusted for the total word count in each teacher's sample (4,344–14,367), they appeared remarkably stable. Words containing /ð/ accounted for 7.4%–11.1% of all words in the teacher talk, which was virtually identical to /ð/ frequency in the 2,500-word BNC corpus (281, or 11.2%).

All /ð/ tokens were accounted for by only 23 unique words at Time 1 and 21 unique words at Time 2. Of these, 19 were identical at both testing times, and all but one (*weather*) came from the 1,000 most frequent word list (Nation, 2006). Even more strikingly, only six word types (*the*, *that*, *this*, *there*, *they*, *then*) accounted for 95% of all /ð/ tokens at Time 1 and 92% at Time 2. The profile of /ð/ was virtually identical in the 2,500-word BNC sample, in which 24 high-frequency word types covered all /ð/ tokens, and only six (*the*, *that*, *they*, *this*, *there*, *their*) accounted for 94% of them.

Although the teacher talk included many repetitions of a few high-frequency words containing /ð/ (e.g., *the*, *that*), this phoneme occurred in different phonetic contexts determined by the immediately preceding segment (e.g., *mother* = vowel, *in the* = nasal). Figure 1 shows the distribution of /ð/ across the eight phonetic contexts, with frequencies expressed as percentage of all /ð/ tokens. Because the three teachers showed identical frequency patterns at both testing times, all data were combined. As Figure 1 illustrates, the largest number of tokens appeared in the sentence or phrase initial context (21%); the smallest number occurred in the voiceless fricative/affricate context (5%). The distribution of /ð/ in the 2,500-word BNC corpus (also shown in Figure 1) follows the same general pattern, with one notable exception: a higher incidence of /ð/ following nasals.

Student Output

The goal of the output analyses was to determine whether student speech reflected the properties of teacher talk. There were 2,453 /ð/ tokens at Time 1 (17.4% of Time 1 word count) and 3,392 tokens at Time 2 (20.9%). As with teacher talk, the use of /ð/ was lexically restricted, with 13 word types used at Time 1 and 17 at Time 2. Six words (*the*, *mother*, *they*, *there*, *brother*, *that*) accounted for 98% of tokens at Time 1, and a similar list (*the*, *they*, *mother*, *there*, *their*, *that*) comprised 99% of tokens at Time 2. The higher percentage of /ð/ in student speech than in teacher talk (17.4%–20.9% vs. 7.4%–11.1%) and some lexical non-overlap between the two reflected task demands. The story characters included the children's mother and two siblings, which would require students to use *mother*, *brother*, *they*, and *their* frequently.

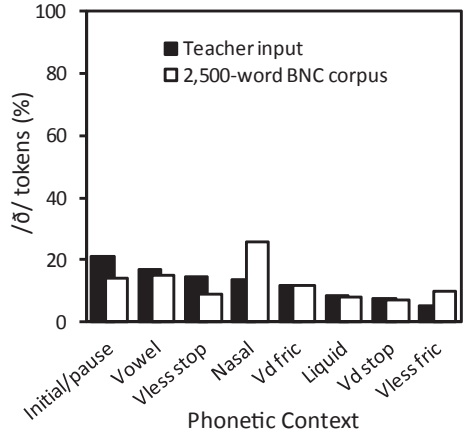

FIGURE 1. Distribution of /ð/ across phonetic contexts in teacher talk and 2,500-word BNC corpus.

Figure 2 shows the distribution of /ð/ across the eight phonetic contexts in student output. Because the distributions were similar at both times, all data were combined. As in the teacher talk, the largest number of tokens appeared in the sentence or phrase initial context (36%); the fewest tokens occurred in the voiceless fricative/affricate context (3%). A comparison of Figures 1 and 2 suggests that the distribution of /ð/ in student output was strikingly similar to the /ð/ profile in the BNC (e.g., both featured a high number of /ð/ following nasals). What distinguished student output from the BNC, however, was a higher incidence of /ð/ preceded by a pause, which is not surprising given that learner speech, compared with proficient speaker output, typically includes pauses and hesitations.

Contrary to our expectations, there was no difference in the accuracy with which students produced /ð/ at Time 1 and Time 2 (26% vs. 23% correct), either when all tokens were combined, $t(119) = 1.40$, $p = .16$, or when each context was considered separately, $t < 1.46$, $p > .15$. This finding was surprising, particularly because the same learners showed sizeable improvements in other L2 skills after similar amounts of instructional input (Collins & White, 2011). However, there was great variability in students' initial accuracy at Time 1 (0%–93% correct), which suggests that students' performance at Time 2 may have been directly related to their initial accuracy at Time 1. Therefore, we divided the students into two equal groups ($n = 60$) based on the median score of production accuracy at Time 1 and compared their accuracy separately. The low group achieved 9% accuracy (0%–18%), compared to the high group's 44% accuracy (19%–93%). Figure 3 shows /ð/ accuracy for these groups at both testing times.

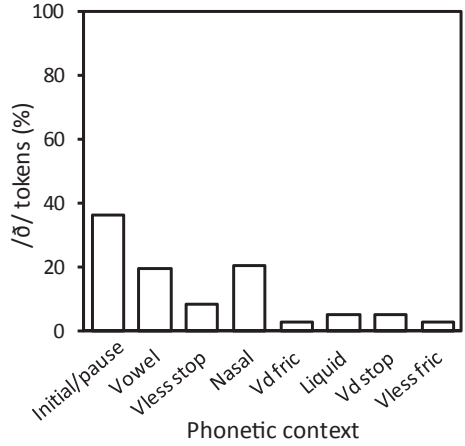

FIGURE 2. Distribution of /ð/ across phonetic contexts in student output.

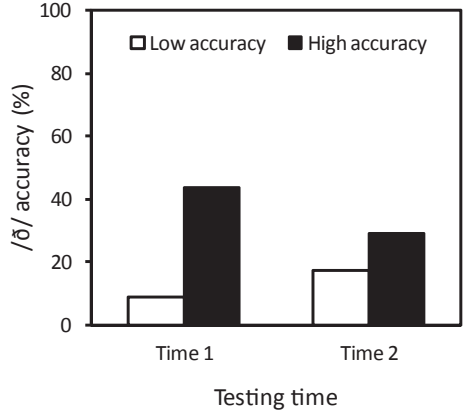

FIGURE 3. /ð/ accuracy for groups with low and high accuracy at Time 1.

A two-way ANOVA with accuracy (low, high) as a between-subjects factor and time (Time 1, Time 2) as a within-subjects factor revealed a significant effect for accuracy, $F(1, 118) = 84.39$, $p < .001$, and a significant accuracy $\cap$ time interaction, $F(1, 118) = 37.70$, $p < .001$, but no significant effect for time. Follow-up comparisons (Bonferroni corrected $\alpha = .0125$) showed that the students in the low group became *more accurate*, attaining 17% correct at Time 2, $t(59) = 3.75$, $p < .001$. However, the high group became less accurate, scoring 29% correct at Time 2, $t(59) = 4.86$, $p < .001$. Nevertheless, there was still a significant difference between the two groups at Time 2, with the

students who were initially more accurate still outperforming those who were initially less accurate, $t(118) = 3.11$, $p = .001$.[3] When we examined students' accuracy separately in each context ($\alpha = .006$), we found that the low group's accuracy gain was restricted to three contexts (initial/pause, nasal, liquid), $t > 2.94$, $p < .005$, whereas the high group's accuracy loss also occurred in only three contexts (initial/pause, nasal, vowel). Notably, most of these contexts (initial/pause, nasal, vowel) were those with the highest /ð/ frequency in teacher talk and student output.

DISCUSSION

Our frequency-based analyses of input and output revealed two findings. We found that the distribution of English /ð/ in student output was similar to the profile of /ð/ in teacher talk and that both matched the frequency properties of a BNC-based sample. In both teacher talk and learner output, /ð/ tokens were restricted to a handful of high-frequency types and varied similarly as a function of the preceding phonetic context. We also found a relationship between learner accuracy and the frequency profiles of teacher talk, such that changes in learners' accuracy over time (gains or losses) were restricted to contextual instances of /ð/ with the highest frequency of occurrence.

Our first finding shows that, with respect to the target investigated here, frequency-based properties of learner speech closely reflect the properties of the language that learners experience around them. In fact, at least in some respects (/ð/ following nasals), learner output was more similar to the profile of "general" English than to the properties of classroom teacher talk. Thus, learners appear to be sensitive to the distributional properties of the input received from the teacher and the language (however minimal) experienced outside the classroom (from television, storybooks, etc.). This implies that frequency patterns in large-scale corpora of general language may provide reasonable estimates of learners' encounters with a language. However, how well such estimates generalize to other features of input (morphology, syntax, other aspects of phonology) in different instructional contexts needs to be investigated further.

Our second finding adds to the growing body of evidence showing that the accuracy of learners' L2 performance reflects input frequencies of the patterns they perceive and produce (e.g., Bradlow & Pisoni, 1999). However, our results qualify this generalization by showing that

[3] Our findings were identical when we removed "outliers" from our analyses (data for eight students whose Time 1 accuracy exceeded 65% correct).

positive effects of frequency in L2 phonological learning may be most pronounced in the earliest stages of learning (when learners' initial accuracy with a given pattern is low). From a theoretical standpoint, this relationship can be described by the power law of practice, which states that effects of frequency on performance are greatest early in the learning and that increased repetition (through either exposure or task performance) after a certain point yields diminishing returns (see DeKeyser, 2007). From a pedagogical viewpoint, this result suggests that providing learners with input containing increased repetitions of targeted phonological features, especially early on, could lead to L2 production benefits.

One intriguing aspect of our results was that the students who were initially more accurate at producing /ð/ became less accurate as the course progressed, which is consistent with U-shaped behavior (i.e., an initial improvement followed by a regression). The idea that increased repetition yields diminishing returns over time cannot fully explain this finding. If frequent exposure to the target in the input has less and less impact on learners' development over time, one would expect learners' development to "fossilize," not necessarily to regress. Clearly, other factors shaped L2 phonological learning in this study. One possibility is that the students were negatively influenced by nontarget /ð/ models in the speech of their peers. Another possibility is that the lexically restricted distribution of /ð/ in the input (featuring a handful of high-frequency tokens like *the* or *that*) did not promote learning beyond high-frequency exemplars after the students had reached a certain level of /ð/ accuracy (see Bybee, 2002). And because the focus in "acquisition-rich" intensive ESL classrooms is on task completion (e.g., telling a story, performing a skit), not on linguistic accuracy, the students never received explicit instruction or corrective feedback on /ð/, which could have "pushed" students to develop further. Some of these factors (e.g., learners' sensitivity to nontarget models in the speech of peers, learners' restricted exposure to lexically diverse input) also rely on frequency characteristics of input to explain L2 development. However, the U-shaped production accuracy points to a possible limitation of frequency-based approaches in explaining *all* aspects of phonological development. Thus, it is also important to take into consideration how learners' existing (and possibly innate) sources of knowledge interact with language input and how explicit focus on language forms and corrective feedback can influence language development.

In terms of pedagogical implications, our findings suggest that learners would benefit from activities in which teachers deliberately focus on the learning targets that occur infrequently in the input. With respect to /ð/, for example, teachers could expose learners to less prototypical examples of /ð/ distributed across a wider range of lexical items (e.g.,

farther, *other*, *without, weather*) in listening activities and then engage learners in focused listening or speaking tasks so that learners could receive feedback on their accuracy. This focus on less prototypical examples could also apply to other learning targets (e.g., simple past, word stress). As we have discussed elsewhere (Collins et al., 2009), a simple way of increasing learners' exposure to learning targets would be for teachers to read aloud to students. Reading aloud of age-appropriate texts may not only increase raw frequencies of targeted items, but also highlight them because teachers speak slowly and with emphasis when reading aloud to students. Taken together, for teachers, our findings illustrate the critical importance of input in instructed L2 settings, and for researchers, they highlight the frequency-based approach as a viable method of investigating input-output relationships in L2 learning.

ACKNOWLEDGMENTS

We are grateful to the two anonymous reviewers for their helpful comments on earlier versions of this article and to the many research assistants who transcribed and assisted with the coding of the original data (reported by Collins et al., 2009). Kathryn MacFadden-Willard, Cassandre McLean Ikauno, Yvette Relkoff, Fabrizio Stendardo, Fernanda Soler, and Derek Theriault provided invaluable assistance with analyzing the teacher input and learner output data reported here. This project was supported by the Social Sciences and Humanities Research Council of Canada and Fonds Québécois de la Recherche sur la Société et la Culture grants awarded to the coauthors.

THE AUTHORS

Pavel Trofimovich is an associate professor of applied linguistics in the Department of Education at Concordia University, in Montreal, Canada. His research focuses on cognitive aspects of second language processing, second language phonology, sociolinguistic aspects of second language acquisition, and the teaching of second language pronunciation.

Laura Collins is an associate professor of applied linguistics in the Department of Education at Concordia University and co-editor of the *Canadian Modern Language Review*. Her research interests include the relationship between pedagogical practices and language learning outcomes, cross-linguistic influence, and the acquisition of tense and aspect.

Walcir Cardoso is an associate professor of applied linguistics in the Department of Education at Concordia University. His research and teaching focus primarily on the acquisition of first and second language phonologies from theoretical and applied perspectives, and on computer-assisted language learning.

Joanna White is an associate professor of applied linguistics in the Department of Education at Concordia University. In her teaching and research, she focuses on

ways of maximizing the benefits of instruction for second language learners across a variety of classroom contexts and language features.

Marlise Horst is an associate professor of applied linguistics in the Department of Education at Concordia University. Her research explores second language vocabulary acquisition and pedagogy from a corpus-informed perspective. She is also interested in the effects of instruction designed to raise learners' cross-linguistic language awareness.

REFERENCES

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074–2085. doi:10.1121/1.427952

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215–221. doi:10.1017/S0272263102002061

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in second language instruction. *TESOL Quarterly, 33*, 655–680. doi:10.2307/3587881

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *Modern Language Journal, 93*, 336–353. doi:10.1111/j.1540-4781.2009.00894.x

Collins, L., & White, J. (2011). An intensive look at intensity in language learning. *TESOL Quarterly, 45*, 106–133. doi:10.5054/tq.2011.240858

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Mahwah, NJ: Lawrence Erlbaum.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188. doi:10.1017/S0272263102002024

Eubank, L., & Gregg, K. R. (2002). News flash—Hume still dead. *Studies in Second Language Acquisition, 24*, 237–247. doi:10.1017/S0272263102002085

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*, 1161–1173. doi:10.1121/1.414884

Gass, S. M., & Mackey, A. (2002). Frequency effects and second language acquisition: A complex picture? *Studies in Second Language Acquisition, 24*, 249–260. doi:10.1017/S0272263102002097

Lightbown, P. M. (1985). Input and acquisition for second language learners in and out of classrooms. *Applied Linguistics, 6*, 263–273. doi:10.1093/applin/6.3.263

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon, England: Multilingual Matters.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and speaking? *Canadian Modern Language Review, 63*, 59–82. doi:10.1353/cml.2006.0049

Piske, T., & Young-Scholten, M. (Eds.). (2009). *Input matters in SLA*. Bristol, England: Multilingual Matters.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York, NY: Routledge.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, England: Cambridge University Press.

Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition, 29,* 407–448. doi:10.1017/S027226310707026X

Extensive Reading Quizzes and Reading Attitudes

TIM STOECKEL
Miyazaki International College
Miyazaki, Japan

NEVITT REAGAN AND FERGUS HANN
Kansai Gaidai University
Hirakata, Japan

doi: 10.1002/tesq.10

■ Extensive reading (ER) has become a common feature of many English as a second or foreign language (ESL/EFL) programs. There is evidence that reading large amounts of easy, interesting material may improve foreign language skills, most notably in vocabulary, reading rates, and overall proficiency. However, teacher evaluation of extensive reading has often been discouraged by ER experts because of its possible negative effect on students' attitudes toward reading. Little investigation has been done to verify this view, and with a substantial increase of available ER assessment materials as well as teachers using them, a gap has grown between expert opinion and current classroom practice.

DEFINITION OF EXTENSIVE READING

Bamford and Day (2004) describe extensive reading as "an approach to language teaching in which learners read a lot of easy material in the new language. They choose their own reading material and read it independently of the teacher. They read for general, overall meaning, and they read for information and enjoyment" (p. 1). Its purpose is "to develop good reading habits, to build up knowledge of vocabulary and structure, and to encourage a liking for reading" (Richards & Schmidt, 2002, pp. 193–194). Typically, students select short readers specifically written at a certain proficiency level or abridged, simplified versions of well-known works of literature. Day (2003) suggests that