ABSTRACT: The clinical importance of clearly identifying variables that are related to across-phoneme generalization and the need for replication of findings that show stimulability as one such variable prompted this study. The first purpose of this study was to replicate the finding that stimulability status of untreated speech sounds predicts generalization patterns, while at the same time examining whether the presence or absence of correct productions could also predict generalization patterns. The second purpose of this study was to replicate the finding that stimulability status of speech sounds targeted in treatment influences the overall amount of across-phoneme generalization. Pretreatment stimulability and percentage correct of misarticulated speech sounds were measured in 7 children with functional phonologic disorders. Following treatment, the relative association between these two variables and generalization to untreated speech sounds was examined. The results indicate that stimulability is the variable that is most strongly associated with generalization. This association between pretreatment stimulability and generalization was present regardless of the stimulability status of the treatment target. Clinical implications of these findings, possible theoretical underpinnings of stimulability, and the usefulness of connected speech samples are discussed.

KEY WORDS: stimulability, phonologic disorders, across-phoneme generalization
was then used as a prognostic indicator for both spontaneous improvement and improvement during treatment. The results of these studies indicated that children with high stimulability scores made significantly greater spontaneous (i.e., without treatment) improvement than did children with low stimulability scores (Carter & Buck, 1958; Farquhar, 1961; Kisatsky, 1967; Snow & Millisen, 1954; Sommers et al., 1967) and that children with low stimulability scores made better progress in treatment than did children with high stimulability scores—possibly the by-product of a ceiling effect (Carter & Buck, 1958; Sommers et al., 1967). However, although these results were statistically significant, not all children conformed to the group trend.

Later studies examined speech sound-specific stimulability (Powell et al., 1991; Powell & Miccio, 1996; Tyler, 1996). In these studies, stimulability was measured similarly to previous studies, but results were presented for individual speech sounds instead of as a general characteristic of the child. From this perspective, a child might be stimulable for one speech sound but not another.

Information regarding speech sound-specific stimulability is typically used in the treatment target selection process. One approach is to select speech sounds that are stimulable as treatment targets so that treatment will be more efficient and the child will experience less frustration (Bleile, 2002; Hodson, Scherz, & Strattman, 2002; Secord, 1989). However, several researchers have recommended targeting sounds that are nonstimulable because those sounds are less likely to improve without direct treatment, and across-phoneme generalization (to stimulable sounds) is more likely to occur if the sound treated is nonstimulable (Gierut, 1998; Miccio, 2002; Miccio et al., 1999; Powell, 1991; Powell et al., 1991).

These findings regarding the role of stimulability in across-phoneme generalization have great potential impact on the treatment of phonologic disorders. If clinicians select nonstimulable speech sounds as treatment targets, treatment will be much more efficient than if they select stimulable sounds. However, it is vitally important that the role of stimulability in across-phoneme generalization be clearly established. Although several studies have been conducted assessing the association between stimulability and spontaneous improvement in articulation (e.g., Carter & Buck, 1958; Farquhar, 1961; Kisatsky, 1967; Snow & Millisen, 1954; Sommers et al., 1967; Tyler, 1996), only two studies have been found that directly assessed the role of stimulability in generalization learning. A total of only 10 children participated in those two studies. Clearly, the findings regarding the role of stimulability in generalization learning need to be replicated to ensure that these findings are not unique to those particular children.

In addition, although both studies demonstrated that untreated stimulable speech sounds are more likely to show generalization than are untreated nonstimulable sounds, only one of the studies (Powell, et al., 1991) demonstrated that the stimulability status of the treatment targets influences across-phoneme generalization. There is an even more pressing need to replicate this finding because it has only been demonstrated in one study. In addition, this conclusion was reached because the 2 participants who were stimulable for the first treatment target did not subsequently generalize to any untreated speech sounds, whereas the 4 participants who were nonstimulable for the first treatment target did generalize to some untreated speech sounds. However, the 2 participants whose first treatment target was stimulable were then taught an additional stimulable speech sound, which led to generalization to three of the four untreated sounds (the three that were stimulable; the remaining sound was non-stimulable). Thus, it is not clear that treatment must occur to a nonstimulable speech sound in order to achieve across-phoneme generalization.

Another reason for replication of the extant findings regarding stimulability is to demonstrate clearly that stimulability is the variable associated with generalization rather than other variables that were not directly examined. For example, Powell et al. (1991) showed that the occasional occurrence of correct production of a misarticulated speech sound during pretreatment measures predicted across-phoneme generalization almost as well as stimulability. Powell et al. found that seven of eight stimulable speech sounds met the generalization criterion whereas only three of eight nonstimulable speech sounds did. However, although all speech sounds measured were absent from the children’s pretreatment phonetic inventories (based on analyses of connected speech and single-word productions collected at the beginning of the study), probe data indicated that four were produced correctly on occasion during baseline (collected after the previously mentioned speech samples, but before treatment). After treatment of one speech sound, all four of these speech sounds met the generalization criterion, whereas only one of the remaining untreated speech sounds that were 0% correct during baseline met the generalization criterion.

Given this closer examination of the data in Powell et al. (1991), it seems possible that pretreatment correct productions may be a variable that is more associated with generalization than stimulability. Although many variables have been examined for their potential relationship to across-phoneme generalization, an exhaustive search of the literature revealed no studies that directly examined pretreatment correct productions as a potential variable. In addition, most studies of this nature examined only a single variable, allowing for the possibility of certain significant variables remaining undiscovered. Because pretreatment correct productions appeared to be related to across-phoneme generalization in at least one study (Powell et al.), a search of the literature was conducted to see if this same relationship would be found (or not found) in other studies. Two studies were found that contained enough information to make this determination. These studies (Dinnsen & Elbert, 1984; Gierut et al., 1987) (a) examined across-phoneme generalization, (b) presented some information regarding pretreatment correct productions, and (c) monitored speech sounds for generalization that were both 0% correct and greater than 0% correct before treatment.

Dinnsen and Elbert (1984) examined pretreatment data for evidence of underlying representations (URs) to see if such data could account for the variability in generalization that was observed. The authors concluded that speech sounds with correct URs were more likely to show...
generalization than sounds with incorrect URs. However, a closer examination of the data yields an alternative explanation. Baseline scores (i.e., occurrences of correct speech sound articulation) for speech sounds with correct URs were higher than baseline scores for sounds judged to have incorrect URs; therefore, the differences in generalization could have been due to differences in pretreatment occurrences of correct articulation rather than to differences in the status of URs.

Gierut et al. (1987) claimed that across-phoneme generalization was more widespread in those children who were taught speech sounds in the order of least to most phonologic knowledge rather than most to least phonologic knowledge. However, generalization tended to occur to speech sounds that were produced correctly at least occasionally during baseline, and generalization tended not to occur to sounds that were 0% correct during baseline (see Figures 2 and 3 [pp. 471–472]).

In these two studies, occasional pretreatment occurrence of correct articulation of misarticulated speech sounds appeared to account for the generalization patterns observed as well as, or even better than, the variable under study. When these findings are combined with the findings of Powell et al. (1991), the potential for a relationship between this variable and across-phoneme generalization seems likely and warrants further investigation.

Research Objectives

The two objectives of the current study were to evaluate (a) the relative association between two pretreatment variables (stimulability and pretreatment percentage correct [Pre%C]) and generalization of correct articulation to untreated speech sounds (GUSS) in children with functional phonologic disorders, and (b) the relationship between treatment target stimulability and GUSS. To accomplish the first objective, stimulability and percentage correct were measured for all misarticulated speech sounds before treatment. Following treatment, the generalization data were examined to determine which variable was unambiguously associated with GUSS, and whether one variable was more powerfully associated with this type of generalization than the other.

The second objective was accomplished by selecting stimulable treatment targets for 3 participants and non-stimulable treatment targets for 4 participants. Differences in GUSS were then compared between these two groups of participants to determine if stimulability of the treatment target appeared to be a functional factor.

METHOD

Participants

Seven children (aged 4:0 [years;months] to 5:7), 6 males and 1 female, fitting the diagnostic classification of children with phonologic disorders of unknown origin (Shriberg, Kwiatkowski, Best, Hengst, & Terselic-Weber, 1986) served as participants in the current study. Children displaying phonologic disorders during a diagnostic evaluation at a university clinic were referred as potential participants for this study. The first 7 children who met the following criteria were selected as participants: (a) percentage of consonants correct (PCC; Shriberg & Kwiatkowski, 1982) less than 85% (mild–moderate to severe phonologic disorder); (b) pure-tone air conduction thresholds within normal limits as measured by a bilateral hearing screening at 20 dB HL (American National Standards Institute, 1970) at 500, 1000, 2000, and 4000 Hz; (c) Type A (normal) tympanogram; (d) a passing score on the Oral Speech Mechanism Screening Examination—Revised (St. Louis & Ruscello, 1987); and (e) no history of mental retardation or other developmental disability. See Table 1 for individual participants’ age, gender, PCC (calculated from the connected speech sample [CSS] described below), corresponding severity adjective, Developmental Sentence Score (DSS; Lee, 1974), and corresponding DSS percentile.

Pretreatment Activities

Data collection. The purpose of pretreatment data collection was to determine which speech sounds were misarticulated by each child and to determine stimulability

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>PCC</th>
<th>Severity</th>
<th>DSS</th>
<th>DSS percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4:0</td>
<td>M</td>
<td>71</td>
<td>Mild–Moderate</td>
<td>5.92</td>
<td>10–25</td>
</tr>
<tr>
<td>B</td>
<td>4:2</td>
<td>M</td>
<td>60</td>
<td>Moderate–Severe</td>
<td>4.28</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>C</td>
<td>5:2</td>
<td>M</td>
<td>49</td>
<td>Severe</td>
<td>4.36</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>D</td>
<td>4:3</td>
<td>M</td>
<td>25</td>
<td>Severe</td>
<td>4.60</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>E</td>
<td>4:2</td>
<td>F</td>
<td>41</td>
<td>Severe</td>
<td>5.02</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>F</td>
<td>5:7</td>
<td>M</td>
<td>33</td>
<td>Severe</td>
<td>5.74</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>G</td>
<td>5:7</td>
<td>M</td>
<td>42</td>
<td>Severe</td>
<td>6.72</td>
<td>10</td>
</tr>
</tbody>
</table>

Note. PCC = Percentage of consonants correct (Shriberg & Kwiatkowski, 1982); DSS = Developmental Sentence Score (Lee, 1974). Age is in years;months.
and Pre%C for those misarticulated speech sounds. Data collection for each participant took place in two phases: collection of a CSS and stimulability testing.

CSS. The CSS was collected in two portions. The first was an unstructured conversation between the child and the first author with various toys present to stimulate verbal behavior. The second portion of the CSS involved the child describing various events taking place in pictures. A significant difference in speech sound production between these two portions of the CSS was not expected (Shriberg & Kwiatkowski, 1985). Pictures for the second portion were selected so that less frequently occurring speech sounds were more likely to be produced in order to supplement the first portion. Toys and pictures were selected such that each phoneme of interest was sampled at least five times in releasing and arresting syllable positions across the two portions of the CSS. Phonemes of interest in the current study were /p, b, t, d, k, g, f, v, s, z, θ, j, tʃ, dʒ, l, r, m, n/ (all the consonants of English except those that rarely or never occur in both the releasing and arresting syllable positions /w, j, h, ð, ʒ, η/). Phonemes in the releasing and arresting positions of syllables were analyzed separately. Therefore, in this study, the term “speech sound” refers to a phoneme of interest in either the releasing or arresting syllable position. A speech sound was considered misarticulated if it was produced incorrectly 50% of the time or more in the CSS.

The number of glossable words in the CSSs ranged from 1024 to 2567 across the participants. All phonologic analyses were based on each participant’s entire CSS.

Stimulability testing. Stimulability testing was conducted at the word level for each misarticulated speech sound in releasing and arresting positions. Each participant produced five words for each misarticulated speech sound preceded by the instructions, “Watch me and listen to me very carefully and try to say exactly what I say” and “Say (word).” The words used for stimulability testing were single syllable words that did not contain consonant clusters or the test phoneme in the other syllable position (e.g., the word “cook” was not used to test for stimulability of /k/). In addition, the five words used for each speech sound contained five different vowels with at least one high, low, front, and back vowel, and the same consonant was not used more than twice in the nonstest syllable position.

Data analyses. A speech sound was labeled stimulable if there was a minimum of one correct production out of five during stimulability testing; otherwise, it was labeled nonstimulable. The Pre%C of each misarticulated speech sound was calculated from the CSS. Misarticulated speech sounds were divided into two groups: those that were produced correctly some of the time (1%–50% correct), and those that were never produced correctly (0% correct). Thus, a minimum of one correct production either during stimulability testing or during the CSS was sufficient for labeling that sound stimulable or occasionally correct.

Data recording. The CSS and stimulability testing for each participant were recorded on a Sony WM D6C tape recorder with an electret condenser tie-clip microphone placed on the shirt approximately six inches from the mouth. All of the participants’ glossable utterances from the CSS were transcribed from the tapes by the first author. Stimulability responses were scored (+/−) on-line.

Treatment

Treatment target selection. Stimulable speech sounds were selected for treatment for 3 participants (A, B, F) and nonstimulable ones for 4 participants (C, D, E, G). All treatment targets except one were speech sounds that were never produced correctly in the CSS (remaining sound = 23% correct). In addition, the speech sound representing the same phoneme as the treatment target, but in the other syllable position, was also produced incorrectly at least 50% of the time in the CSS. Each child was taught one or two speech sounds, resulting in a total of 10 treatment targets across participants. For participants with two treatment targets, stimulability status was the same for both (i.e., both stimulable or both nonstimulable).

Treatment program. Treatment was conducted three times per week for 30–50-min sessions. All treatment sessions were conducted by the first author. The treatment used was a series of steps beginning with imitated production of the target speech sound in isolation and ending with production of the target speech sound in words in self-generated sentences. Several studies examining generalization have also followed a similar treatment format (Bowen & Cupples, 1999; Costello & Onstine, 1976; Elbert & McReynolds, 1978, 1985; Gierut, 1998; Hoffman, 1983; McReynolds & Bennett, 1972; Morrisette & Gierut, 2002; Powell, Miccio, Elbert, Brasseur, & Strike-Roussos, 1999; Rvachew & Nowak, 2001). Small modifications of the general format were occasionally made so as to maximize learning for each participant, in keeping with the principles of programmed instruction (Costello, 1977). For the 3 participants who had two treatment targets, treatment of each speech sound was conducted serially.

Treatment for a particular speech sound was terminated when pass criterion (90% [18/20]) was met on the last treatment step, which required correct production of the target speech sound in self-generated sentences. However, if generalization of correct articulation of the target speech sound to untreated words occurred before this, treatment was terminated at that time.

All responses during treatment were immediately scored as correct or incorrect. All treatment sessions were recorded using the same equipment that was used for pretreatment data collection.

Generalization probes. In order to assess generalization of correct articulation to untreated words and to untreated speech sounds, generalization probes were constructed for each participant. Because generalization probe words were to be evoked spontaneously through picture naming, an effort was made to include words that were already in the participants’ vocabularies. To accomplish this, lexical probes, wherein children were asked to identify pictures, were administered before the construction of the generalization probes, which were then constructed using, whenever possible, words that the participants had correctly identified.
The generalization probes consisted of words selected to assess generalization of correct articulation (a) of the target speech sound to untreated words, and (b) to untreated speech sounds (GUSS). To assess generalization to untreated words, a minimum of five words containing the treatment target was included in the generalization probe. Generalization was considered to have occurred if correct production occurred in at least 80% of the untreated words on at least one generalization probe, without the occurrence of overgeneralization. To assess GUSS, a minimum of five words for each of a participant’s untreated misarticulated speech sounds was included in the generalization probe (see below for generalization criterion). The number of items on the generalization probes ranged from 76 to 130 across the participants.

The generalization probes were administered three times before the initiation of treatment (to establish baselines), periodically throughout treatment, and three times following the end of treatment. For those participants who had two treatment targets, generalization probes were administered a minimum of three times after treatment on the first target was terminated and before treatment was initiated on the second target. All responses to the generalization probes were transcribed and recorded using the same procedures used for pretreatment data.

Treatment words. Treatment was initiated for each treatment target using five words. For some treatment targets, more words were added later in the treatment program to encourage generalization to untreated words if it was not occurring (Elbert, Powell, & Swartzlander, 1991). Treatment words were selected from the lexical probes after the generalization probes had been constructed. When possible, treatment words were selected that had been correctly identified on the lexical probes. None of the words in the generalization probe was used in treatment.

Reliability

Interjudge percentage agreement was calculated for all speech data collected: CSSs, stimulability testing, treatment data, and generalization probes. All reliability observations were based on audiotapes of the relevant sessions. The primary observer was the first author. Two reliability observers were used: (a) A graduate student in speech-language pathology with training in phonetic transcription served as the reliability observer for stimulability testing and treatment data, and (b) the second author served as the reliability observer for the CSSs and generalization probes. The student displayed intrajudge agreement of at least 80% for each of the two types of data recording she would perform.

Pretreatment data.

CSSs. The number of utterances for which percentage agreement was calculated and the total number of utterances in the CSS for each of three randomly selected participants are as follows: Participant A: 259/344 (75%); Participant C: 511/881 (58%); Participant G: 484/601 (81%). Interjudge percentage agreement was calculated for those speech sounds that were (a) speech sounds of interest in the current study, and (b) transcribed as incorrect by at least one of the observers. An agreement was considered to have occurred when the transcription of both observers reflected an incorrect production. Percentage agreement was calculated only for incorrect productions because misarticulated speech sounds were the focus of the study and so that the resulting percentage agreement would not be artificially inflated by the more frequent correctly produced sounds. The item-by-item percentage agreement formula used for all reliability analyses was (# Agreements/Agreements + Disagreements) × 100. Item-by-item percentage agreement for each of the CSSs for each of the three participants was as follows: Participant A: 60%; Participant C: 80%; Participant G: 76%.

Stimulability testing. Interjudge percentage agreement was calculated for the entire stimulability test for three other randomly selected participants. Both the primary observer and the reliability observer recorded each response as “+” (target speech sound produced correctly) or “−” (target speech sound produced incorrectly). Item-by-item percentage agreement for stimulability testing for the three participants was as follows: Participant B: 74%; Participant F: 91%; Participant G: 84%.

Treatment data. Interjudge percentage agreement was calculated for a randomly selected 10% of the treatment sessions for each participant. Both the primary observer/clinician and the reliability observer recorded each response as “+” (correct) or “−” (incorrect). The primary observer “paused” the tape recorder during all response-contingent feedback to the participants during treatment so that the reliability observer was able to make independent judgments from the audiotape recordings. In addition, the treatment sessions selected for reliability assessment were presented in a random order to the reliability observer so that she was unaware of the actual temporal sequence of the treatment sessions. Trial-by-trial percentage agreement for responses during treatment sessions for all participants ranged from 70% to 100% (median = 94%).

Generalization probes. Interjudge percentage agreement for each participant was calculated for two randomly selected probes administered in pre- and post-nontreatment conditions and for a randomly selected 10% of all probes administered during treatment.

Item-by-item percentage agreement was calculated for all misarticulated speech sounds that were being directly tested. The reliability observer transcribed all speech sounds in each word and was unaware of the particular speech sounds for which percentage agreement would be calculated. An agreement was considered to have occurred if the transcription of both observers reflected a correct production or an incorrect production. Probe numbers were removed from the data sheets and tapes, and the probes were randomized for the reliability observer so that she was unaware of the actual temporal sequence of the generalization probes. Item-by-item percentage agreement

Because percentage agreement for Participant A’s spontaneous speech sample was low (60%), all analyses affected by this sample of speech were reperformed using data from the reliability observer. In no case did the analyses of these data change the original interpretations. Results of these reanalyses are available from the first author.
stimulable speech sounds will not (95 of 108). The sounds will show generalization (20 of 24) and non-sounds matched the hypothesis that stimulable speech showed generalization. These data indicate that stimulable speech sounds are likely to show generalization if they are stimulable and are unlikely to do so if they are not. Conversely, only 12% of speech sounds that were non-stimulable showed generalization. These data indicate that speech sounds are likely to show generalization if they are stimulable and are unlikely to do so if they are not. The strength of the association between stimulability and GUSS was quantified in two ways: (a) hypothesis testing and (b) calculation of a correlation coefficient. Eighty-seven percent (115) of the 132 untreated speech sounds monitored, 24 were stimulable and 108 were not. For stimulable speech sounds, 83% showed generalization. The first objective of the current study was to examine the relationship between stimulability and Pre%C and GUSS. GUSS was measured for all untreated speech sounds that were misarticulated before treatment (<50% correct). A speech sound was considered to have shown generalization if that speech sound was produced correctly in at least 20% more of the probe items than in any baseline probe, in at least three probes administered after baseline. Across all 7 participants, there were 132 untreated misarticulated speech sounds included in these analyses. The percentage of untreated speech sounds that showed generalization as a function of stimulability is shown in the first column of Table 2. Of the 132 untreated speech sounds monitored, 24 were stimulable and 108 were not. For stimulable speech sounds, 83% showed generalization. Conversely, only 12% of speech sounds that were non-stimulable showed generalization. These data indicate that speech sounds are likely to show generalization if they are stimulable and are unlikely to do so if they are not. The percentage of untreated speech sounds that showed generalization as a function of pretreatment stimulability (stimulable, nonstimulable), Pre%C in connected speech (1–50%, 0%), and Pre%C in single words (1–50%, 0%) is shown in Table 2. Of the 132 untreated speech sounds monitored, 24 were stimulable and 108 were not. For stimulable speech sounds, 83% showed generalization. The second column of Table 2 displays generalization as a function of Pre%C and illustrates that this variable is also associated with GUSS. Slightly less than half (40%) of the 52 speech sounds that were 1%–50% correct showed generalization; only 15% of the 80 speech sounds that were 0% correct showed generalization. From the former group, speech sounds that showed generalization ranged in actual levels of correct production from 1%–42% correct, essentially the same as the 1%–43% correct range across speech sounds that did not display generalization. As was done with stimulability, the strength of the association between Pre%C and GUSS was quantified with hypothesis testing and calculation of a correlation coefficient. Sixty-seven percent (89) of the 132 untreated speech sounds matched the hypothesis that speech sounds produced correctly on occasion (1%–50% correct) will show generalization, and speech sounds that are never produced correctly (0%) will not show generalization. The correlation of \( r = .286 \) that was obtained between Pre%C and GUSS, although lower in shared variance than the correlation obtained between stimulability and GUSS, was found to be statistically significant (\( p < .001 \)). Because the correlations between both variables and GUSS were statistically significant, a common criterion test (Hotelling, 1940) was conducted to determine if one variable was more strongly associated with GUSS than the other. The results of this test indicated that the relationship between stimulability and GUSS is stronger than the relationship between Pre%C and GUSS (\( t = 4.631, p < .001 \)). Table 3 shows, for each participant, the percentage of untreated misarticulated speech sounds that matched the hypothesis for each variable examined above. Although the sample sizes are too small for statistical analyses, it can be seen that the individual data for all 7 participants followed the group trend—that stimulability was more strongly associated with GUSS than Pre%C. Because the criterion used to label a sound “stimulable” was set low (only one correct production out of five opportunities), the stimulability testing data were additionally examined to determine the appropriateness of that criterion. Of the 24 speech sounds determined to be stimulable, 10 were produced correctly only once during
stimulability testing. The remaining 14 speech sounds were fairly evenly distributed between two, three, four, and five correct productions (between two and six speech sounds per level of accuracy). Across all 24 stimulable speech sounds, 83% (20) showed generalization. The number of stimulable speech sounds that were produced correctly only once during stimulability testing that showed generalization was 80% (8 out of 10), thus supporting the use of the criterion used in this study.

The Relationship Between Treatment Target Stimulability and GUSS

The second objective of the current study was to examine the relationship between stimulability status of treatment targets and GUSS. The percentage of untreated speech sounds showing generalization for participants with stimulable treatment targets is shown in the first column of Table 4. Of the 45 untreated speech sounds that were monitored, 13 were stimulable and 32 were not. For stimulable speech sounds, 85% showed generalization. Conversely, only 16% of speech sounds that were nonstimulable showed generalization. Across the 3 participants whose treatment targets were stimulable, 84% (38) of the speech sounds matched the hypothesis that stimulable speech sounds will show generalization (11 of 13) and nonstimulable speech sounds will not (27 of 32).

The percentage of untreated speech sounds showing generalization for participants with nonstimulable treatment targets is shown in the second column of Table 4. Of the 87 untreated speech sounds that were monitored, 11 were stimulable and 76 were not. For stimulable speech sounds, 82% showed generalization. Conversely, only 11% of speech sounds that were nonstimulable showed generalization. Across the 4 participants whose treatment targets were nonstimulable, 89% (77) of the speech sounds matched the hypothesis that stimulable speech sounds will show generalization (9 of 11) and nonstimulable speech sounds will not (68 of 76).

Table 3. Percentage (and number) of untreated misarticulated speech sounds that matched the hypothesis for each variable (stimulability & Pre%C) for each participant and for all participants.

<table>
<thead>
<tr>
<th>Participant (number of misarticulated speech sounds)</th>
<th>Stimulability</th>
<th>Pretreatment percentage correct in connected speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (11)</td>
<td>82% (9)</td>
<td>45% (5)</td>
</tr>
<tr>
<td>B (10)</td>
<td>90% (9)</td>
<td>70% (7)</td>
</tr>
<tr>
<td>C (18)</td>
<td>89% (16)</td>
<td>83% (15)</td>
</tr>
<tr>
<td>D (29)</td>
<td>97% (28)</td>
<td>62% (18)</td>
</tr>
<tr>
<td>E (19)</td>
<td>84% (16)</td>
<td>68% (13)</td>
</tr>
<tr>
<td>F (24)</td>
<td>83% (20)</td>
<td>71% (17)</td>
</tr>
<tr>
<td>G (21)</td>
<td>81% (17)</td>
<td>67% (14)</td>
</tr>
<tr>
<td>All part. (132)</td>
<td>87% (115)</td>
<td>67% (89)</td>
</tr>
</tbody>
</table>

Table 4. Percentage (and number) of untreated speech sounds to which generalization of correct articulation occurred as a function of pretreatment stimulability (stimulable, nonstimulable) in participants whose treatment targets were stimulable and nonstimulable.

<table>
<thead>
<tr>
<th>Stimulable treatment targets</th>
<th>Nonstimulable treatment targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stim (13)</td>
<td>Not stim (32)</td>
</tr>
<tr>
<td>% Generalized</td>
<td>85</td>
</tr>
<tr>
<td># Generalized</td>
<td>11</td>
</tr>
<tr>
<td>Stim (11)</td>
<td>Not stim (76)</td>
</tr>
<tr>
<td>% Generalized</td>
<td>82</td>
</tr>
<tr>
<td># Generalized</td>
<td>9</td>
</tr>
</tbody>
</table>

Across all 7 participants, 87% of the 132 untreated speech sounds matched the hypothesis that stimulable speech sounds will show generalization and nonstimulable speech sounds will not. For the 3 participants whose treatment targets were stimulable, 84% of the 45 untreated speech sounds matched this hypothesis. For the 4 participants whose treatment targets were nonstimulable, 89% of the 87 untreated speech sounds matched this hypothesis. Thus, when untreated speech sounds were separated by stimulability status of the treatment target, a notable difference in GUSS was not found.

DISCUSSION

Two variables were examined for their respective associations with GUSS. Using the percentage of misarticulated speech sounds that matched each hypothesis and statistical analyses as gauges of the strength of the association between each variable and GUSS, stimulability was the variable that was most powerfully associated with generalization (83% of the untreated stimulable speech sounds showed generalization). Pre%C was not as strongly associated with generalization (40% of the speech sounds that were 1%–50% correct showed generalization). This supports the findings of Miccio et al. (1999) and Powell et al. (1991) that stimulable speech sounds are more likely to show generalization than sounds that are nonstimulable. However, the second finding of Powell et al., that across-phoneme generalization did not occur when treatment targets were stimulable, was not supported. In the current study, generalization of correct articulation to stimulable speech sounds occurred with essentially the same frequency in participants whose treatment targets were either stimulable or nonstimulable.

Although the conclusions drawn from this study and from the one by Powell et al. (1991) regarding the influence of the stimulability status of treatment targets on GUSS differ, the results of the two studies do not actually conflict. In Powell et al., only 2 participants were taught stimulable treatment targets. The authors based their conclusions on the fact that no generalization across speech sounds occurred in those 2 participants. However, the
analyses were conducted after very little treatment. In fact, the 4 participants whose treatment targets were non-stimulable also demonstrated very little generalization (only five speech sounds out of the 16 monitored across the 4 participants). For the 2 participants whose treatment targets were stimulable, treatment was also conducted on a second stimulable speech sound. Following this treatment, those 2 participants generalized to all untreated stimulable speech sounds, displaying the same generalization pattern as the other 4 participants who were taught nonstimulable speech sounds. So although the two studies draw different conclusions, the results appear to be similar.

**Stimulability Versus Pretreatment Percentage Correct**

One of the interesting findings of the current study was the lack of correspondence between a speech sound’s stimulability and its Pre%C. Across all 7 participants, there were 52 misarticulated speech sounds that were produced correctly on occasion (1%–50% correct) in extensive CSSs (more than double the number of stimulable speech sounds). However, almost two thirds of these (33 [63%]) were not produced correctly during stimulability testing (i.e., were nonstimulable). Even though they were occasionally correct in connected speech (as high as 24% correct), 28 of the 33 showed no generalization. Indeed, of the 21 speech sounds that were occasionally correct in connected speech and also showed generalization, 76% (16) were also stimulable. In contrast, of the 31 speech sounds that were occasionally correct in connected speech but did not show generalization, only 10% (3) were also stimulable. Therefore, for predicting GUSS, a speech sound’s stimulability status is again shown to outweigh the influence of its appearance in connected speech.

There may be other explanations for at least some of the sporadic correct productions of generally misarticulated speech sounds that occurred in the CSSs. Perhaps some were products of coarticulation wherein correct productions were peculiar to their particular phonetic contexts. As well, some could have been the product of listener misperceptions spawned by a difficult listening task. Supporting this explanation are the reliability findings of this study: 60%, 76%, and 80% agreement for speech sounds judged in connected speech as compared to 74%, 84%, and 91% agreement on speech sounds in single-word stimulability measures and a median 90% level of agreement across all participants found for the single-word generalization probes.

Logically, the presence of occasional correct productions of a speech sound in spontaneous speech implies the motoric ability to produce that speech sound. If stimulability testing also measures motoric ability to produce a speech sound, then a high correspondence between these two measures might be expected. However, this was not the case, and similar findings have been reported elsewhere (Lof, 1996). Although a child must have the motoric ability to produce a speech sound in order to produce it correctly during stimulability testing, such testing (at least when conducted at the word level as it was in the current study) must also measure some variable(s) beyond motor skill. Some researchers have proposed that producing a speech sound during word-level stimulability testing requires lexical knowledge (i.e., a correct UR) of that speech sound (Dinnsen & Elbert, 1984; Elbert & McReynolds, 1978). If true, then stimulability could be viewed as a measure of URs. Other possibilities are that stimulability might measure variables such as metalinguistic awareness, metaphonetic awareness, or the existence of an underlying speech-motor program. Regardless of which additional variable(s) is being measured, the motor ability to produce a speech sound in spontaneous connected speech is apparently not sufficient for being able to produce that speech sound during stimulability testing.

**Pretreatment Percentage Correct Revisited**

The hypothesis that Pre%C is significantly related to the occurrence of GUSS resulted from a close examination of three studies (Dinnsen & Elbert, 1984; Gierut et al., 1987; Powell et al., 1991). In these three studies, the occasional pretreatment occurrence of correct production of misarticulated speech sounds was primarily or entirely determined from single-word productions. However, in the current study, Pre%C was determined from connected speech samples. Research has indicated that many children correctly produce sounds in single words that they misarticulate in connected speech (e.g., DuBois & Bernthal, 1978; Morrison & Shriberg, 1992; Shriberg & Kwiatkowski, 1985). For comparison purposes, the data from this study were reanalyzed to determine the nature of the relationship between Pre%C in single words and GUSS. Pre%C in single words was determined for each sound by dividing the total number of correct productions in the three baseline probes by the total number of opportunities to produce the sound (minimum of 15 opportunities per sound).

Although the above-mentioned research would appear to predict that some sounds never produced correctly in connected speech would be produced correctly on occasion in single words, this was only true of two of the 80 relevant untreated speech sounds. Conversely, 37 of the 52 speech sounds that were produced correctly occasionally in connected speech were never produced correctly in single words. Thus, overall, misarticulated speech sounds were produced less accurately in single words than in connected speech for these participants.

When examining the association between Pre%C in single words with GUSS, 65% of the 17 speech sounds that were 1%–50% correct showed generalization, whereas only 19% of the 115 speech sounds that were 0% correct showed generalization (see the third column of Table 2). Hypothesis testing revealed that 79% (104) of the speech sounds matched the hypothesis that speech sounds produced correctly on occasion (1%–50% correct) in single words will show generalization, and speech sounds that are never produced correctly (0%) in single words will not show generalization. The correlation of $r = .353$ that was obtained between Pre%C in single words and GUSS is statistically significant ($p < .001$). However, whereas Pre%C in single words appears to be related to GUSS more
strongly than Pre%C in connected speech (79% vs. 67% of the speech sounds matched the hypothesis, respectively), a common criterion test revealed that this difference is not statistically significant ($t = 0.744, p > .05$). (And a common criterion test comparing the relative association between Pre%C in single words and stimulability with GUSS was statistically significant [$t = 3.69, p < .001$].)

Although this reanalysis of Pre%C data does not change the initial findings of this study (stimulability is still a stronger predictor of GUSS than Pre%C, whether measured in connected speech or in single words), it does question the long-held view that the collection of connected speech samples is necessary to fully capture children’s phonologic errors. Not only did Pre%C in single words predict GUSS for more untreated speech sounds than did Pre%C in connected speech (although not at a statistically significant level), the single-word samples identified as “mis-articulated” every speech sound that was identified as such in connected speech. In other words, if the single-word speech samples collected in this study had been used exclusively, there would have been no difference in the identification of misarticulated speech sounds. (Recall that this required a minimum of five familiar words per sound.)

Given the significant amount of time required to collect and analyze connected speech samples, this may not be the most efficient use of clinicians’ time. On a cautionary note, however, how misarticulated speech sounds were produced in single words versus connected speech was not analyzed in this study. It is possible that the exclusive use of single-word samples could lead to different conclusions regarding patterns of misarticulations present.

**Implications for Treatment and Suggestions for Further Research**

The results of the current study have implications for treatment, primarily in the area of treatment target selection. In order to make treatment efficient, it is desirable to achieve the maximum amount of generalization possible. The results of the current study indicate that stimulable speech sounds are more likely to show generalization than nonstimulable speech sounds. This would seem to indicate that stimulable speech sounds should have lower priority as treatment targets because they are likely to improve as a function of generalization when other speech sounds are treated. This concept of selecting treatment targets that are in some way more complex (i.e., incorrect UR, non-stimulable, later developing, low frequency) is not new, although it may be contrary to common practice. Several studies have suggested that selecting treatment targets in this way is likely to increase generalization across phonemes (Gierut, 2001; Gierut et al., 1987; Gierut et al., 1996; Morrisette & Gierut, 2002; Powell et al., 1991).

Although a rationale for this phenomenon has not been directly advanced, one possibility is that the prerequisite skills necessary for production of less complex aspects of the phonologic system are necessarily learned during the acquisition of more complex aspects.

One possible drawback to the selection of nonstimulable treatment targets is the time required for treatment. In this study, treatment of nonstimulable speech sounds required an average of 50% more sessions than treatment of stimulable sounds. Further research is needed to determine if longer treatment time for nonstimulable sounds is a common phenomenon. Presumably, treatment of nonstimulable speech sounds would still be more efficient than treatment of stimulable sounds, even if it takes longer to teach nonstimulable sounds, due to across-phoneme generalization, but this needs to be determined empirically.

Because stimulability was so strongly related to generalization across phonemes in the current study, this variable should be examined for its potential relationship to other kinds of generalization such as generalization of correct articulation to spontaneous speech and across settings, because enhancing these types of generalization would also make treatment more efficient.

Another issue that should be investigated in further research is the method of determining stimulability. In the current study, stimulability was assessed at the word level. An interesting question would be whether stimulability testing conducted in nonsense syllables or with speech sounds in isolation would also be associated with certain types of generalization. This needs to be determined empirically.

The clinical value of determining factors that are likely to enhance generalization cannot be understated. Making treatment more efficient results in earlier dismissal, which results in smaller caseloads for clinicians. Therefore, research that identifies factors related to generalization helps not only children with phonologic disorders, but other children receiving services from speech-language pathologists working in the public schools as well.

**ACKNOWLEDGMENTS**

This study reports, in part, the results of a dissertation that was completed by the first author. We are grateful to the children and their families for their participation in this research. We would like to thank Susan Higgins, a graduate student at Ball State University, who served as a reliability observer.

**REFERENCES**


therapy for kindergarten, first and second grade children. 

*Journal of Speech and Hearing Research, 10, 428–437.*


Contact author: Cari B. Flint, Department of Communication Disorders, California State University at Los Angeles, 5151 State University Drive, Los Angeles, CA 90032. E-mail: cflint@calstatela.edu