Purpose: The purpose of this narrative review was to examine the evidence for imitation training for targeting grammar in children with developmental language disorder. Method: Studies investigating imitation training were compiled from two databases—PsycINFO and Linguistics and Language Behavior Abstracts—as well as from reference lists of several books and articles reviewing language therapy. Results: Twenty-one studies were reviewed. We summarized the demographic and methodological characteristics of the included studies. We identified substantial differences in the implementation of imitation training across the studies and noted a lack of studies considering how those differences might affect therapy outcome. Studies showed a rapid increase in the production of grammatical targets during imitation training activities, higher performance relative to a control group that did not receive any therapy, and generalization to untrained exemplars. However, the studies did not unequivocally link the increased performance to the imitation training therapy. Studies also showed limited generalization to conversational speech. Conversationally based therapy resulted in faster or higher usage in conversational speech, while imitation training resulted in faster immediate gains in usage during therapy trials. Conclusion: We believe that imitation training can continue to be a useful tool in a speech-language pathologist’s toolbox, but with suggested modifications. Specifically, we suggest use of imitation training as a means for rapidly achieving production of previously absent grammatical targets. However, we do not recommend sole or long-term use of imitation training for working on grammar. Further research is needed to refine our use of the procedure.

An important clinical issue for speech-language pathologists (SLPs) is to use therapy approaches that are supported by evidence. Specifically, SLPs want evidence that a particular therapy approach will lead to the largest amount of change in the shortest amount of time. One highly popular procedure used by SLPs is elicited imitation—the provision of a prompt for a child to replicate an utterance modeled by the clinician. In a recent survey by Finestack and Satterlund (2018), 98% of practicing SLPs providing early intervention (i.e., serving children ages 0–5 years) reported using requests for elicited imitation (73% frequently and 25% sometimes), while 0% reported never using this intervention technique. SLPs serving elementary school children responded similarly, with 95% reporting use of elicited imitation (61% frequently and 34% sometimes) and only 2% reporting that they never used this technique. Clearly, SLPs who serve children rely on elicited imitation as a teaching strategy. It is, therefore, incumbent upon our field to examine the evidence for elicited imitation and determine its place in evidence-based practice. It was also reported in the survey that 83% of SLPs in early intervention and 91% of SLPs in elementary education used drill activities either sometimes or frequently. Considering these findings together, it seemed likely that elicited imitation is often used within drill activities. In this report, we specifically examine the evidence for the use of elicited imitation within the context of highly structured drill activities, a procedure termed “imitation training.”

Most scholars agree that children do not learn language by imitation alone. That is, imitation as the only learning mechanism is not sufficient to account for the myriad of language skills that a typically developing child acquires. Yet, the ability to imitate has been shown to be an important skill for cognitive and social development (Meltzoff & Williamson, 2013; Tomasello, 1999). Furthermore, children who are more imitative end up with better language. For example, Bates et al. found that an infant’s ability to imitate predicted later spoken language ability.
This effect is seen for older children as well. A study by Speidel and Herreshoff (1989) found that first-grade children who spontaneously imitated more frequently were able to produce longer utterances and that the effects of imitation were significantly more pronounced in the longest category of utterances.

Other researchers have questioned the role of imitation in language learning. In a study of preschool children by Bloom et al. (1974), imitation was limited to structures that children had already acquired or that were produced in the immediately following session. This raises the possibility that imitation might not contribute to language learning if children imitate only things that they already know. Consistent with this possibility and in contrast with Speidel and Herreshoff (1989), Tager-Flusberg and Calkins (1990) found nonimitative utterances to be longer and more advanced than imitative utterances in the spontaneous speech of 2-year-old children.

It must be noted, however, that self-selected or spontaneous imitation—children choosing to imitate an adult’s preceding utterance of their own accord—is vastly different from elicited imitation—children imitating an utterance because they have been explicitly instructed to do so. The use of such elicited imitation in therapy has received support in the literature. Fey and Proctor-Williams (2000) recommend the use of elicited imitation as an intervention technique, particularly when first introducing new forms, because it provides practice in production that may strengthen and stabilize syntactic representations. Indeed, imitative responses, when explicitly requested, can be grammatically more advanced than a child’s spontaneous utterances (Bredin-Oja & Fey, 2014; Fraser et al., 1963). Additionally, elicited imitation serves to dramatically increase the frequency of producing target structures compared to production rates found in a language learning environment without prompts to imitate and highlights grammatical forms assumed not to be readily perceived by children with developmental language disorder (DLD; Fey & Proctor-Williams, 2000).

Paul and Norbury (2012) note that it may be an advantage that imitation training does not replicate natural language learning. These authors suggest that, because they have not learned language under natural circumstances, children with DLD might need a less natural learning context. The highly structured nature of imitation training enables clinicians to provide condensed input and elicit a large number of responses in a short amount of time, yielding a high dosage of practice trials. Thus, Paul and Norbury recommended using imitation training as a first step in language therapy to establish forms that are absent or infrequently used and then following up with a less structured and more embedded approach to achieve generalization to conversational speech.

The purpose of this review is to examine the evidence for imitation training for targeting grammar in children with DLD. We have reviewed a number of published studies and summarized the, albeit limited and dated, evidence here. Given the quality of studies, we chose to do a narrative review that is broader in scope instead of a systematic review that typically emphasizes randomized controlled studies. The restricted focus of a systematic review has a limited capacity to guide clinical practice because not all the available evidence is considered (MacLure, 2005). Narrative reviews, on the other hand, are comprehensive summaries combined with interpretation and critique and, therefore, can more fully describe the nature of the knowledge base that informs clinical practice (Greenhalgh et al., 2018).

**Terminology for Using Models in Therapy**

We used the term “model” to refer to any exemplar of a target form presented by the clinician, regardless of whether or not the child is asked to attempt production of the target after hearing the model. While the use of models is an essential component of all language therapies, how models are used varies widely across therapy approaches. The terminology for characterizing the different ways in which models are used in therapy has varied across studies, making it difficult to distinguish between therapy approaches and compare results. Before reviewing the studies, we will, therefore, first suggest a uniform terminology that we will use in our review.

**Elicited Production and Imitation**

In “elicited production,” the clinician uses prompts to encourage the child to attempt the target form. We limited use of this term to instances in which the child is prompted to independently generate the target form. In contrast, we used “elicited imitation” when the child is prompted to replicate a clinician model of the target. Figure 1 illustrates these elicitation procedures. Both elicited production and elicited imitation can be used in the repetitive format of structured drill activities or within embedded communicative interactions to elicit individual responses.

**Antecedent and Consequent Models**

“Antecedent models” are given before the child’s production attempt. That is, the clinician provides a model and prompts the child to replicate that model (i.e., elicited imitation). “Consequent models” are given after the child’s production. That is, the clinician prompts the child to generate the target form (i.e., elicited production) and

Figure 1. Comparison of elicited production and elicited imitation.

<table>
<thead>
<tr>
<th>a. Elicited production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician prompt for production with a question or comment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b. Elicited imitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician model of the target utterance &amp; prompt for production with a question or command</td>
</tr>
</tbody>
</table>
subsequently models the target. Consequent models that follow incorrect child productions are also referred to as “corrective recasts or expansions” (e.g., Camarata & Nelson, 1992; Goldstein, 1984; Paul & Norbury, 2012). Note though that consequent models can also be given after correct child productions (e.g., Valian & Casey, 2003). After providing a consequent model, the clinician could prompt the child to replicate the modeled utterance (i.e., elicited imitation). Figure 2 illustrates these modeling options.

Imitation Training

We used the term “imitation training” to refer to structured drill activities that involve a series of models and elicited imitations of those models (see Figure 3). Imitation training trials include presentation of a picture or other nonverbal stimulus, a model of the target form, and a prompt to elicit imitation of the model (Eisenberg, 2013; Fey & Proctor-Williams, 2000). Observational modeling can be implemented by having the child just listen to the models without subsequently attempting production of the target utterances (e.g., Connell, 1987), or it can be followed by elicited production—prompting the child to talk about the same or a different set of pictures without hearing further models (e.g., Courtright & Courtright, 1976, 1979). Figure 3 illustrates both of these implementation options for observational modeling.

Embedded Therapy

Our review also includes several studies in which imitation training was compared to a different technique, which we referred to as “embedded therapy.” In this procedure, the child listens to a set of models of the target form while looking at pictures or watching the clinician use doll figures to act out the utterances (Eisenberg, 2013; Fey & Proctor-Williams, 2000). Observational modeling can be implemented by having the child just listen to the models without subsequently attempting production of the target utterances (e.g., Connell, 1987), or it can be followed by elicited production—prompting the child to talk about the same or a different set of pictures without hearing further models (e.g., Courtright & Courtright, 1976, 1979). Figure 3 illustrates both of these implementation options for observational modeling.

Method

PsycINFO and Linguistics and Language Behavior Abstracts were used to search the literature for research studies about imitation training. We used broad search terms to locate reports that used different terminologies to refer to this therapy approach, although this was likely to result in a larger number of irrelevant studies to be manually reviewed. Key word search terms included “child,” “language,” and “imitation” combined with “intervention,” “therapy,” or “treatment.” We limited the search to peer-reviewed publications written in English. No limits were placed on publication dates, although the search was completed in December 2018. Reference lists from several books and articles reviewing language therapy were also consulted (i.e., Ebbels, 2014; Eisenberg, 2013; Fey, 1986; Leonard, 2014).

The title, abstract, and then the method section of each research study were examined to determine whether the article was a research study that provided evidence about the efficacy of imitation training for targeting grammar goals. This included studies of invented morphemes and syntactic forms. A study was accepted as imitation training if it involved a structured drill activity in which a clinician provided a model and elicited an imitative response from the child. This included studies that compared imitation training to another therapy. The population of interest was children with DLD, but we also accepted studies of children with typical language (TL) as we had no a priori reason to expect that children with TL would respond differently than children with DLD.

Studies were eliminated from the review for the following reasons: (a) The study was about a topic other than therapy (e.g., assessment, acquisition), (b) the study focused on aspects of language other than grammar (e.g., vocabulary, pragmatics), (c) the study focused exclusively on children with language disorders associated with another syndrome (e.g., autism, intellectual disability, hearing impairment), or (d) the study focused on therapy or instruction for children who were second-language learners. Studies were further restricted based on levels of evidence (LOEs) from the American Academy for Cerebral Palsy and Developmental Medicine (Darrah et al., 2008), listed in Table 1. To be consistent with the broad scope of a narrative review, only studies involving expert opinion and case reports (Level 5) were excluded. All included studies thus had some degree of experimental control.

Results

The database searches yielded 3,400 references after eliminating duplicates. Of these, only 13 studies met the inclusion criteria. An additional eight studies were identified through the reference lists, bringing the total of included studies to 21. Most of the eliminated studies either did not focus on therapy or did not focus on grammar goals. Seventeen studies were excluded because the participants included only children with language disorders associated with autism, intellectual disability, or hearing impairment. Two studies of imitation training were eliminated because they were single-case studies (Level 5). One other group study was eliminated because the therapy did not meet our definition of imitation training. Although a model was provided before each child response, the clinician did not prompt the child to replicate the model.

1Note that Fey and Proctor-Williams referred to this as “elicited imitation.”
We first summarized demographic and methodologic characteristics of the included studies. We next reviewed how imitation training was implemented across the studies. Studies were divided into three groups: studies investigating only imitation training, studies comparing imitation training to observational modeling, and studies comparing imitation training to embedded therapies. Within each of these groupings, we separately reviewed studies that used antecedent models and studies that used consequent models.

**Demographic and Methodological Characteristics**

**Source and Publication Date**

The included studies were published between 1972 and 2003 in 11 different journals, with 52% of the studies published in 2003.

**Figure 2. Comparison of antecedent and consequent models.**

- a. Antecedent model (before child production attempt)

- b. Consequent model (after child production attempt)

- c. Consequent corrective model followed by elicited imitation

**Figure 3. Comparison of imitation training and observational modeling.**

- a. Imitation training (each elicited imitation item includes a clinician model/prompt and a child imitative response)

- b. Observational modeling

- c. Observational modeling followed with elicited production (each elicited production item includes a clinician prompt and a child nonimitative response)
Table 1. Levels of evidence.

<table>
<thead>
<tr>
<th>Level</th>
<th>Group studies</th>
<th>Single-subject design studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Large randomized controlled trials (N ≥ 100)</td>
<td>Randomized controlled N-of-1 study involving an ATD or a concurrent MBD with staggered initiation of the therapy across ≥ 3 participants or behaviors</td>
</tr>
<tr>
<td>2</td>
<td>Smaller randomized controlled trials (N &lt; 100)</td>
<td>Nonrandomized, controlled, concurrent MBD across ≥ 3 participants or behaviors</td>
</tr>
<tr>
<td>3</td>
<td>Cohort study with concurrent control group</td>
<td>Nonrandomized, nonconcurrent, controlled MBD with ≥ 3 participants or behaviors</td>
</tr>
<tr>
<td>4</td>
<td>Cohort study without concurrent control group Case series</td>
<td>Nonrandomized, controlled, single-subject design with ≥ 3 phases (ABA, ABAB, BAB, etc.)</td>
</tr>
<tr>
<td>5</td>
<td>Retrospective case reports Expert opinion</td>
<td>Single-case study with AB design</td>
</tr>
</tbody>
</table>

Note. Adapted with permission from the American Academy for Cerebral Palsy and Developmental Medicine (AACPDM; Darrah et al., 2008). ATD = alternating or simultaneous treatment design; MBD = multiple baseline design.

aRandomization of participants to condition. bRandomization of participants or treatment targets to condition. cConcurrent initiation of baseline sessions for participants or targets with staggered initiation of therapy sessions.

Participants

Table 3 shows the number of participants for group and single-subject studies. There were 15 group studies ranging from eight to 80 participants, with most (87%) having fewer than 50 participants and four (27%) having 10 or fewer participants. The six single-subject studies ranged from one to six participants, with three (50%) having only one or two participants. Note that three studies (Camarata & Nelson, 1992; Camarata et al., 1994; Nelson et al., 1996) were classified as group rather than single-subject studies because the data were pooled across participants and mean performance for each condition was reported as for a group design.

Participant age (see Tables 6–8) ranged from 2;2 to 10;0 (years;months), with all but one study including participants between the ages of 3;0 and 6;10. Seventeen studies included only children with DLD, two studies included only children with TL, and four studies included both children with DLD and TL.

Therapy Targets

Therapy targets are included in Tables 6–8. Most of the studies (73%) targeted the same forms for all participants, including sentence constituents (2), subject case pronouns (2), possessive morpheme ‘s (1), use of auxiliary and/or copula BE (5), and questions with auxiliary and/or copula BE (2). In five of these studies, the targets were invented grammatical forms, including noun suffixes (2), verb suffixes (1), and word ordering rules (2). One study assigned different targets to each participant—either copula BE or auxiliary BE. Five other studies individualized targets based on language sample analysis and probes and included noun and verb morphemes, sentence types, and complex sentence forms as the therapy targets.

In 12 (57%) of the studies, all targets were completely absent (i.e., 0% usage) at baseline. One study included only partially mastered targets (i.e., up to 65% usage) at baseline. Five (24%) of the studies included both absent and partially mastered targets. The remaining three studies did not specify the baseline level of the target forms.

Service Delivery

Service delivery is shown in Table 4. In most (90%) of the studies, therapy was provided individually in a clinic. As can be seen from the table, scheduling and dosage varied considerably.

LOE

Table 5 summarizes the LOE across the included studies. Two thirds of the group studies were classified as Level 2. However, although meeting the randomization criteria for this level, six of the studies were very small. Three of these studies, namely, Camarata and Nelson (1992), Haynes and Haynes (1980), and Mulac and Tomlinson (1977), included 10 or fewer total participants, and three other studies, namely,
Anderson (2001), Courtright and Courtright (1979), and Matheny and Panagos (1978), had group sizes of less than 10. The remainder of the group studies (33%) was classified as Level 3 cohort studies without randomized assignment of participants to condition. Most of the single-subject studies (83%) were classified as either Level 1 or 2. However, although meeting the design criteria for these levels, two of the studies, namely, Hegde et al. (1979) and Hughes and Carpenter (1989), included only one or two participants. Given these limitations on sample size, only seven of the included studies would have met the criteria for a systematic review. The classification of each study is included in Tables 6–8.

**Implementation of Imitation Training**

Imitation training involves the provision of clinician models and elicited imitation of those models within a highly structured drill activity. All of the studies involved an exact match between the model and the child’s response. That is, the clinician provided a model of the target form, and the child was asked to repeat the model verbatim (exact imitation, also called “mimicry”; Haynes & Haynes, 1980). There was otherwise considerable variation across the studies in how imitation training was implemented. We summarize the differences here. Additional details are shown in Table 9.

1. **Timing of the prompt relative to the child’s response**: Figure 5 illustrates this variable. In most of the reviewed studies, the clinician provided a prompt before saying the model so that the child’s response occurred immediately after the model (immediate imitation, e.g., *Say the boy is running*; Connell, 1987; Connell & Stone, 1992; Hegde, 1980; Hegde & Gierut, 1979; Hegde et al., 1979; Hughes & Carpenter, 1989; Matheny & Panagos, 1978; Mulac & Tomlinson, 1977; Nelson et al., 1996; Valian & Casey, 2003). Alternatively, in some of the studies, the clinician first modeled the target and then prompted the child to imitate (delayed imitation, e.g., *The boy is running. You say that*; Anderson, 2001; Courtright & Courtright, 1976, 1979; Gillum et al., 2003). Timing of the prompt was not specified in the other seven studies.

2. **Timing of the model relative to the child’s response** (see Figure 2): In most of the reviewed studies, models were provided before child responses (antecedent models; Anderson, 2001; Camarata & Nelson, 1992; Camarata et al., 1994; Cole & Dale, 1986; Connell, 1987; Connell & Stone, 1992; Courtright & Courtright, 1976, 1979; Friedman & Friedman, 1980; Gillum et al., 2003; Haynes & Haynes, 1980; Hegde, 1980; Hegde & Gierut, 1979; Hegde et al., 1979; Hughes & Carpenter, 1989; Matheny & Panagos, 1978; Mulac & Tomlinson, 1977; Nelson et al., 1996; Valian & Casey, 2003; Zwitman & Sonderman, 1979). Alternatively, in two of the studies, models were provided after child responses (consequent models; Goldstein, 1984; Valian & Casey, 2003).

3. **Contingency of models**: For most of the reviewed studies with antecedent models, the child heard a model before each attempt regardless of response accuracy (Anderson, 2001; Camarata & Nelson, 1992; Camarata et al., 1994; Cole & Dale, 1986; Connell, 1987; Connell & Stone, 1992; Courtright & Courtright, 1976, 1979; Friedman & Friedman, 1980; Gillum et al., 2003; Haynes & Haynes, 1980; Hughes & Carpenter, 1989; Matheny & Panagos, 1978; Mulac & Tomlinson, 1977; Nelson et al., 1996; Valian & Casey, 2003; Zwitman & Sonderman, 1979). In some of these studies, the provision of antecedent models was abruptly discontinued after the child achieved some accuracy criterion such that therapy shifted from elicited imitation to elicited production (Camarata & Nelson, 1992; Camarata et al., 1994; Hughes & Carpenter, 1989; Nelson et al., 1996). In several studies, however, antecedent models were contingent on the child’s accuracy. Hegde and colleagues discontinued antecedent models after four consecutive correct responses and then reinstated antecedent models for four subsequent...
trials after the child made an error on a nonimitative production (Hegde, 1980; Hegde & Gierut, 1979; Hegde et al., 1979). In one study with only consequent models, a model was provided after each incorrect response, and the child was prompted to repeat the model (termed a “corrective model”; Goldstein, 1984). These options are illustrated in Figure 6.

4. Density of models: In most studies, models were provided with a consistently high density, that is, before every attempt by the child (Camarata & Nelson, 1992; Camarata et al., 1994; Cole & Dale, 1986; Courtright & Courtright, 1976, 1979; Friedman & Friedman, 1980; Gillum et al., 2003; Haynes & Haynes, 1980; Hughes & Carpenter, 1989; Matheny & Panagos, 1978; Mulac & Tomlinson, 1977; Nelson et al., 1996; Valian & Casey, 2003; Zwitman & Sonderman, 1979). A few studies provided an even higher density of models, either by providing two antecedent models of the target sentence before each attempt (Connell, 1987; Connell & Stone, 1992; Nelson et al., 1996; Valian & Casey, 2003) or by also providing a consequent model after the child’s attempt, regardless of correctness (Valian & Casey, 2003). An alternative would be to present models less frequently, for instance, before every two or three child responses, so that imitative attempts alternated with nonimitative trials. This occurred in the study by Anderson (2001) because they alternated therapy trials with test trials that were not preceded by a model. For still other studies in which models were provided only when the child made errors, density varied as a function of child accuracy (Goldstein, 1984; Hegde, 1980; Hegde & Gierut, 1979; Hegde et al., 1979). This also resulted in a lower overall density of models.

5. Consequences after attempts: Most of the studies provided some form of reinforcement after correct imitations. The most common reinforcer was praise or social reinforcement (e.g., “good”). In many of the other studies, this was combined with either a primary reinforcer (food) or tokens.

We found only one study that compared different ways of implementing imitation training. The study by Valian and Casey (2003) compared outcome for imitation training with differing density and/or timing of models. This study included three conditions: one in which imitation training was implemented with a single antecedent model before child attempts (see Figure 7a); a second condition in which the child heard two antecedent models before each attempt (see Figure 7b); and a third condition with both an antecedent model and a consequent model, with the latter provided after the child’s production regardless of the correctness of the child’s imitation (see Figure 7c). Inversion for trained auxiliaries (BE, can) increased in all three conditions. However, generalization of inversion to untrained auxiliaries (will, DO) occurred only for the two conditions in which the child heard two models, with no difference between these two conditions. We otherwise lack studies comparing these differences in implementation.

Table 4. Service delivery in the included studies.

<table>
<thead>
<tr>
<th>General</th>
<th>n (%)</th>
<th>Session lengtha</th>
<th>n (%)</th>
<th>Sessions per weeka,b</th>
<th>n (%)</th>
<th>Total sessions</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>19 (90)</td>
<td>10–19 min</td>
<td>1 (5)</td>
<td>1 (5)</td>
<td>1 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small group</td>
<td>2 (9)</td>
<td>20–29 min</td>
<td>6 (29)</td>
<td>5 (25)</td>
<td>3–6</td>
<td>7 (33)</td>
<td></td>
</tr>
<tr>
<td>30–39 min</td>
<td>3 (14)</td>
<td>3 (15)</td>
<td>3 (15)</td>
<td>4 (19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–49 min</td>
<td>3 (14)</td>
<td>4 (10)</td>
<td>2 (10)</td>
<td>3 (14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 min</td>
<td>1 (5)</td>
<td>5 (25)</td>
<td>1 (5)</td>
<td>5 (15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 min</td>
<td>1 (5)</td>
<td>NS</td>
<td>8 (40)</td>
<td>≥125</td>
<td>2 (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120 min</td>
<td>1 (5)</td>
<td>NS</td>
<td>5 (24)</td>
<td>CNT</td>
<td>3 (14)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. NS = not specified; CNT = could not tell.
aThe lower value was used for studies with variable session length or sessions per week. 
bThe total number sums to only 20 because one study involved a single session.

Table 5. Level of evidence of the included studies.

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Number (% of studies)</th>
<th>Group</th>
<th>Single subject</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2 (33)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10 (67)</td>
<td>2 (33)</td>
<td>12 (57)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5 (33)</td>
<td>0</td>
<td>5 (23)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2 (33)</td>
<td>2 (10)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>6</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

Eisenberg et al.: Use of Imitation Training for Targeting Grammar 211
Table 6. Studies of imitation training (IT).

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Therapy targets</th>
<th>Baseline</th>
<th>Therapy</th>
<th>Delivery and dose</th>
<th>Level of evidence</th>
<th>Results</th>
<th>Maintenance</th>
<th>Generalize to untrained exemplars</th>
<th>Generalize to speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hegde (1980)</td>
<td>DLD 5:6–5:10</td>
<td>S1: aux is, S2: cop is</td>
<td>Absent</td>
<td>Immediate exact imitation</td>
<td>1:1 3 days/wk, 40–45 min</td>
<td>Level 4 SSD</td>
<td>S1 (aux): A1 = increase to 90% usage; B = drop to 30% usage; A2 = increase to 100% usage</td>
<td>Not measured</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hegde et al. (1979)</td>
<td>DLD 3:9–4:0</td>
<td>Subject pronouns (he, she, they) and aux are</td>
<td>Absent</td>
<td>Immediate exact imitation</td>
<td>1:1 4–5 days/wk, 45 min</td>
<td>Level 2 SSD</td>
<td>S1 (aux): A1 = increase to 96% usage; B = drop to 23% usage; A2 = increase to 100% (aux 1st): no change in baseline until therapy began for aux ‘s, aux was, and poss</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
<tr>
<td>Hegde &amp; Gierut (1979)</td>
<td>DLD 4:9</td>
<td>Not specified</td>
<td>Absent</td>
<td>Imitation</td>
<td>1:1; 17 30-min sessions for 4 wks</td>
<td>Level 2 SSD</td>
<td>Reduction in number of trials to achieve criterion (10 correct) on each target sentence</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
<tr>
<td>Matheny &amp; Panagos (1978)</td>
<td>DLD 5:5–6:10</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Immediate exact imitation</td>
<td>1:1; 5 months</td>
<td>Level 2 group IT &gt; cA for syntax, but both made significant gains in syntax; cN made no significant gains in syntax</td>
<td>Not measured</td>
<td>Not measured</td>
<td>Not measured</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Therapy targets</td>
<td>Baseline</td>
<td>Therapy</td>
<td>Delivery and dose</td>
<td>Level of evidence</td>
<td>Results</td>
<td>Generalize to untrained exemplars</td>
<td>Generalize to speech</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>----------</td>
<td>---------</td>
<td>------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>Mulac &amp; Tomlinson (1977)</td>
<td>DLD</td>
<td>cop and aux is questions</td>
<td>Absent</td>
<td>Immediate exact imitation; IT+ added transfer training</td>
<td>1:1; 120 min/wk (60 min, 2 days/wk; 30 min, 4 days/wk)</td>
<td>Level 2 group</td>
<td>Both IT and IT+ achieved 100% usage in therapy sessions; did not report outcome for cA but not significantly different from outcome for therapy groups</td>
<td>Yes, 3 wks after IT+ training concluded</td>
<td>Not measured</td>
<td>No, for IT</td>
</tr>
<tr>
<td>Valian &amp; Casey (2003)</td>
<td>DLD</td>
<td>wh-questions</td>
<td>Partially mastered</td>
<td>Exact imitation</td>
<td>1:1; 4 days/wk</td>
<td>Level 3 group</td>
<td>Increase target use in all three conditions; IT1 &lt; IT2 and IT3</td>
<td>Not measured</td>
<td>No for IT1; yes for IT2 and IT3 with IT2 = IT3</td>
<td>Not measured</td>
</tr>
<tr>
<td>Zwitman &amp; Sonderman (1979)</td>
<td>N-is-Ving-N</td>
<td>Emerging/partially mastered</td>
<td>Emerging/partially mastered</td>
<td>Exact imitation</td>
<td>1:1; 1 day/wk, 45 min</td>
<td>Level 3 group</td>
<td>IT &gt; cN for mean gain in usage of targets, but five in cN made substantial gains</td>
<td>Not measured</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Note. In baseline, absent = 0% usage, emerging = ≤ 10% usage, and partially mastered = 11%–65% usage. DLD = developmental language disorder; aux = auxiliary; cop = copula; 1:1 = one-to-one individual therapy; wk(s) = week(s); SSD = single-subject design; poss = possessive; cN = no therapy control; cA = articulation therapy control; IT+ = IT with extended transfer training; IT1 = one antecedent model; IT2 = two antecedent models; IT3 = one antecedent model and one consequent model.
Table 7. Studies comparing imitation training (IT) and observational modeling (OM).

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Therapy targets</th>
<th>Baseline</th>
<th>Comparison therapy</th>
<th>Delivery and dose</th>
<th>Level of evidence</th>
<th>Results</th>
<th>Maintenance</th>
<th>Generalize to untrained exemplars</th>
<th>Generalize to connected speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson (2001)</td>
<td>DLD and TL</td>
<td>Invented verb suffix</td>
<td>Absent</td>
<td>Production of nontarget utterance after each model</td>
<td>1:1; one session with 32 trials each</td>
<td>Level 2 group</td>
<td>IT = OM for both DLD and TL</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
<tr>
<td>Connell (1987)</td>
<td>DLD and TL</td>
<td>Invented noun suffix</td>
<td>Absent</td>
<td>Just listening</td>
<td>1:1; three 20-min sessions with 20 trials each</td>
<td>Level 2 group</td>
<td>DLD: IT &gt; OM for correct responses to probes; TL: IT &lt; OM for correct responses to probes</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
<tr>
<td>Connell &amp; Stone (1992)</td>
<td>DLD and TL</td>
<td>Invented noun suffixes</td>
<td>Absent</td>
<td>Production of nontarget utterance after each model</td>
<td>1:1; four sessions in total, two sessions of 20 trials each</td>
<td>Level 3 group</td>
<td>TL: IT = OM DLD: IT &gt; OM for rate of correct responses to probes</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
<tr>
<td>Courtright &amp; Courtright (1976)</td>
<td>DLD</td>
<td>Subject pronoun they</td>
<td>IT: emerging/partially mastered OM: absent</td>
<td>Attempted same utterances after hearing sets of 10 utterances</td>
<td>1:1; three 20-min sessions with 20 trials each</td>
<td>Level 3 group</td>
<td>Performance increased in both IT and OM conditions First probe: IT &gt; OM Last probe: IT &lt; OM IT: plateau after second session OM: continued increase across all sessions</td>
<td>Higher accuracy after 1 week in modeling condition</td>
<td>IT &lt; OM (i.e., higher accuracy in OM on probe for untrained exemplars)</td>
<td>Not measured</td>
</tr>
</tbody>
</table>
Table 7. (Continued).

### a: Studies of imitation training with antecedent models

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Therapy targets</th>
<th>Baseline</th>
<th>Comparison therapy</th>
<th>Delivery and dose</th>
<th>Level of evidence</th>
<th>Results</th>
<th>Maintenance</th>
<th>Generalize to untrained exemplars</th>
<th>Generalize to connected speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courtright &amp; Courtright (1979)</td>
<td>DLD 3:11–6;11 #IT = 12 #OM = 12</td>
<td>Invented grammatical form</td>
<td>Absent</td>
<td>Attempted novel utterances after hearing sets of 10 utterances</td>
<td>1:1; three 20-min sessions with 20 trials each</td>
<td>Level 2 group</td>
<td>IT &lt; OM for accuracy on probe after each session</td>
<td>Higher accuracy after 1 week in modeling condition</td>
<td>IT &lt; OM (i.e., higher accuracy in OM on probe for untrained exemplars)</td>
<td>Not measured</td>
</tr>
<tr>
<td>Haynes &amp; Haynes (1980)</td>
<td>DLD 4 years #IT = 5 #OM = 5</td>
<td>Copula /s</td>
<td>Absent/ emerging</td>
<td>Attempted novel utterances after hearing sets of 10 utterances</td>
<td>1:1; five sessions with 40 trials each</td>
<td>Level 2 group</td>
<td>IT &gt; OM for rate of correct responding for all five sessions</td>
<td>No difference on probe after 1 week</td>
<td>IT = OM (i.e., not significantly different overall); however, over the four probes, plateau for IT versus consistent increase and higher final level for OM</td>
<td>Not measured</td>
</tr>
</tbody>
</table>

### b: Studies of imitation training with consequent models

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Therapy targets</th>
<th>Baseline</th>
<th>Comparison therapy</th>
<th>Delivery and dose</th>
<th>Level of evidence</th>
<th>Results</th>
<th>Maintenance</th>
<th>Generalize to untrained exemplars</th>
<th>Generalize to connected speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldstein (1984)</td>
<td>TL 4:11–5;7 #IT = 6 Low SES</td>
<td>Nonsense four- to six-word combinations (of previously learned two-word combinations)</td>
<td>Absent</td>
<td>Just listening</td>
<td>1:1; 3–5 days/ week for 20–25 min</td>
<td>Level 1 SSD</td>
<td>IT &gt; OM for all six children</td>
<td>Not measured</td>
<td>Yes</td>
<td>Not measured</td>
</tr>
</tbody>
</table>

**Note.** In baseline, absent = 0% usage, emerging = ≤ 10% usage, and partially mastered = 11%–65% usage. DLD = developmental language disorder; TL = typical language; SES = socioeconomic status; 1:1 = one-to-one individual therapy; SSD = single-subject design.
### Table 8. Studies comparing imitation training (IT) and conversational therapy (C).

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Therapy targets</th>
<th>Baseline</th>
<th>Comparison therapy</th>
<th>Delivery and dose</th>
<th>Level of evidence</th>
<th>Results</th>
<th>Maintenance</th>
<th>Generalize to untrained exemplars</th>
<th>Generalize to connected speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camarata &amp; Nelson (1992)</td>
<td>DLD 4;9–5;11 (years; months)</td>
<td>Varied morphemes and syntactic forms</td>
<td>Absent</td>
<td>CRT</td>
<td>1:1 30 min, 2 days/wk for 32 days</td>
<td>Level 2 group</td>
<td>IT = C for number of presentations prior to achieve first elicited production; IT &gt; C for number of presentations to achieve first spontaneous use</td>
<td>Not measured</td>
<td>Not measured</td>
<td>IT &lt; C for mean number of nontargeted productions</td>
</tr>
<tr>
<td>Camarata et al. (1994)</td>
<td>DLD 4;0–6;10</td>
<td>Varied morphemes and syntactic forms</td>
<td>Absent</td>
<td>CRT</td>
<td>1:1 25 min, 2 days/wk for 24 days</td>
<td>Level 2 group</td>
<td>IT &lt; C for number of clinician presentations and number of sessions to achieve first elicited production; IT &gt; C for number of clinician presentations and number of sessions to achieve first spontaneous production</td>
<td>Not measured</td>
<td>Not measured</td>
<td>IT &lt; C for total number of nontargeted productions</td>
</tr>
<tr>
<td>Cole &amp; Dale (1986)</td>
<td>DLD 3;2–5;9</td>
<td>Not specified</td>
<td>Not specified</td>
<td>“Interactive language instruction” similar to CRT</td>
<td>4:1 2-hr classes, 2 days/wk for IT and 3 days/wk for C for 32 wks</td>
<td>Level 2 group</td>
<td>IT &lt; C for DSS, PLS, and NSST Neither IQ nor language score predicted gain from either therapy</td>
<td>Not measured</td>
<td>Not measured</td>
<td>Not measured</td>
</tr>
<tr>
<td>Friedman &amp; Friedman (1980)</td>
<td>DLD 3;2–5;9</td>
<td>Varied morphemes and syntactic forms</td>
<td>Not specified</td>
<td>ILDT</td>
<td>2:1 in class 1 hr, 4 days/wk for 8 mos</td>
<td>Level 3 group</td>
<td>For gains in DSS: Higher IQ: IT &lt; C Lower IQ: IT &gt; C LL: IT &lt; C, with large gains for IT and moderate gains for C HL: IT &lt; C, with little gains for IT and moderate gains for C</td>
<td>Not measured</td>
<td>Not measured</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*(table continues)*
| Study          | Participants          | Therapy targets                                      | Baseline | Comparison therapy | Delivery and dose | Level of evidence | Results | Maintenance          | Generalize to | Generalize to         |
|---------------|-----------------------|-----------------------------------------------------|----------|--------------------|-------------------|-------------------|---------|----------------------| untrained      | connected speech     |
| Gillum et al. (2003) | DLD 4;3–6;8, n = 4 | Poor imitation skills | Absent | CRT                | 1:1 25 min, 2 days/wk for 24 days | Level 1 SSD | IT < C for number of trials to criterion (criterion = 5 elicited productions) | Not measured | Not measured | Not measured |
| Hughes & Carpenter (1989) | Copula /is and are, n = 4 | Absent/partially mastered | 1:1 30 min, 3 days/wk for 4 wks | ILDT | Level 4 SSD | IT > C for three of the participants | Not measured | IT = C | Little or no generalization to at-home conversation for either condition |
| Nelson et al. (1996) | DLD and TL 2;2–6;7, #DLD = 7, #TL = 7 | Varied morphemes and syntactic forms; six per child | Three absent and three partially mastered | CRT | Level 2 group | Absent targets: IT > C for number of presentations to achieve first spontaneous use for both DLD and TL; three with DLD did not achieve spontaneous production in IT | Not measured | Not measured | IT < C for number of targets generalized to home use (31% vs. 77%) |

Note. In baseline, absent = 0% usage and partially mastered = up to 65% usage. DLD = developmental language disorder; CRT = conversational recast therapy (Baker & Nelson, 1984); 1:1 = one-to-one individual therapy; wk(s) = week(s); 4:1 = small group with four children; DSS = Developmental Sentence Score; PLS = Preschool Language Scale; NSST = Northwestern Screening Syntax Test; ILDT = interactive language development teaching (Lee, 1974); mos = months; LL = low language; HL = high language; SSD = single-subject design; TL = typical language.
Table 9. Implementation of imitation training.

<table>
<thead>
<tr>
<th>Study</th>
<th>Match between response and model</th>
<th>Timing of response after prompt</th>
<th>Timing of the model</th>
<th>Contingency of models</th>
<th>Density of models</th>
<th>Consequence after attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson (2001)</td>
<td>Exact</td>
<td>Delayed</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt); test trials alternated with therapy trials</td>
<td>Not specified</td>
</tr>
<tr>
<td>Camarata &amp; Nelson (1992)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>Model discontinued after 90% accuracy</td>
<td>1:1 (before every attempt)</td>
<td>c: praise and token reinforcement</td>
</tr>
<tr>
<td>Camarata et al. (1994)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>Model discontinued after 90% accuracy</td>
<td>1:1 (before every attempt)</td>
<td>c: praise and/or token reinforcement</td>
</tr>
<tr>
<td>Cole &amp; Dale (1986)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt)</td>
<td>c: praise</td>
</tr>
<tr>
<td>Connell (1987)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>NA</td>
<td>2:1 (two models before every attempt)</td>
<td>c: correction request</td>
</tr>
<tr>
<td>Connell &amp; Stone (1992)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>NA</td>
<td>2:1 (two models before every attempt)</td>
<td>c: consequating event (an animated character acts out the child's utterance)</td>
</tr>
<tr>
<td>Courtright &amp; Courtright (1976)</td>
<td>Exact</td>
<td>Delayed</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt)</td>
<td>No reinforcement</td>
</tr>
<tr>
<td>Courtright &amp; Courtright (1979)</td>
<td>Exact</td>
<td>Delayed</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt)</td>
<td>No reinforcement</td>
</tr>
<tr>
<td>Friedman &amp; Friedman (1980)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>Contingency for fading the model not specified</td>
<td>1:1 (before every attempt)</td>
<td>c: not specified</td>
</tr>
<tr>
<td>Gillum et al. (2003)</td>
<td>Exact</td>
<td>Delayed</td>
<td>Ant</td>
<td>Model discontinued after 90% accuracy</td>
<td>1:1 (before every attempt)</td>
<td>c: praise and/or token reinforcement</td>
</tr>
<tr>
<td>Goldstein (1984)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Cons</td>
<td>Model after each incorrect response before the second attempt</td>
<td>Varied based on child's accuracy</td>
<td>c: token</td>
</tr>
<tr>
<td>Haynes &amp; Haynes (1980)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt)</td>
<td>c: social reinforcement</td>
</tr>
<tr>
<td>Hegde (1980)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>Model discontinued after four correct and reinstated after two incorrect trials</td>
<td>Varied based on child's accuracy</td>
<td>c: candy and verbal reinforcement</td>
</tr>
<tr>
<td>Hegde &amp; Gierut (1979)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>Models initiated on the next trial after error; discontinued after four correct imitative trials</td>
<td>Varied based on child's accuracy</td>
<td>c: candy and verbal reinforcement</td>
</tr>
<tr>
<td>Hegde et al. (1979)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>Model discontinued after four correct and reinstated after two incorrect trials</td>
<td>Varied based on child's accuracy</td>
<td>c: “no” and averting eye contact</td>
</tr>
<tr>
<td>Hughes &amp; Carpenter (1989)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>Model discontinued after three sets of five correct responses</td>
<td>1:1 (before every attempt)</td>
<td>c: token and verbal praise</td>
</tr>
<tr>
<td>Matheny &amp; Panagos (1978)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt); test trials alternated with therapy trials</td>
<td>c: not specified</td>
</tr>
<tr>
<td>Mulac &amp; Tomlinson (1977)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>NA</td>
<td>1:1 (before every attempt)</td>
<td>c: not specified</td>
</tr>
</tbody>
</table>

(table continues)
task as the imitation group. Similarly, in the study by Mulac and Tomlinson (1977), the gain in performance for the imitation training condition was not significantly different from the gain made by a control group of children who had received articulation therapy. In addition, although Zwitman and Sonderman (1979) reported larger gains in mean production accuracy for the imitation training group, five children in the no-therapy control group showed as large an increase in production as children who had received imitation training.

Several single-subject design studies by Hegde and colleagues (Hegde, 1980; Hegde & Gierut, 1979; Hegde et al., 1979) examined the efficacy of imitation training using contingent antecedent models after errors (illustrated in Figure 6c). Hegde and Gierut (1979) reported an increase in production for untrained and trained exemplars of the target form for each of the two participants. Hegde et al. (1979) conducted a single-subject study with staggered multiple baselines across four or five morpheme targets for two children. Subject 1 showed no change from baseline until after the initiation of therapy for auxiliary *is*, auxiliary *was*, and possessive *–s*. However, there was an increase in production of copula *is* after therapy for the auxiliary. Similarly, Subject 2 showed no change from baseline until after the initiation of therapy for copula *is*, copula *was*, and possessive *–s*, but there was an increase in the production of auxiliary *is* and *was* after therapy for the comparable form of the copula. Hegde (1980) also found a significant crossover in production accuracy between copula and auxiliary. The reason for this crossover effect is unclear. Hegde concluded that auxiliary *BE* and copula *BE* represented a single response class to the children given the similarity in their surface form. If so, then the crossover could reflect a desirable generalization to an untrained exemplar.

A frequently mentioned limitation of imitation training is a lack of generalization to untrained exemplars or conversational speech (Fey, 1986; Paul & Norbury, 2012). Two studies (Hegde & Gierut, 1979; Zwitman & Sonderman, 1979) reported generalization to untrained exemplars (i.e., use of the targeted forms in different linguistic contexts). Hegde (1980) reported high levels of usage in conversational speech for copula and auxiliary once the child achieved 100% usage on therapy trials: ≥ 90% usage in conversation with the clinician in the clinic and 89% usage with the parent at home. However, Mulac and Tomlinson (1977) found no generalization to conversation without adding a procedure for transfer training with the parent at home. Another type of generalization is to untrained members of the target category. In the study by Valian and Casey (2003), there was no generalization of question inversion to untrained auxiliaries when participants heard a single antecedent model before the elicited imitation attempt (per Figure 7a). However, generalization to untrained auxiliaries did occur when modeling density was increased so that participants either heard two antecedent models (per Figure 7b) or heard a model both before and after the elicited imitations (per Figure 7c).

### Table 9. (Continued).

<table>
<thead>
<tr>
<th>Study</th>
<th>Match between response and model</th>
<th>Timing of response after prompt</th>
<th>Timing of the model</th>
<th>Contingency of models</th>
<th>Density of models</th>
<th>Consequence after attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson et al. (1996)</td>
<td>Exact</td>
<td>Immediate</td>
<td>Ant</td>
<td>Model discontinued after 90% accuracy</td>
<td>Two models before every attempt</td>
<td>c: verbal praise and tokens that could be traded for a prize</td>
</tr>
<tr>
<td>Valian &amp; Casey (2003)</td>
<td>Exact</td>
<td>Immediate</td>
<td>C1: Ant</td>
<td>NA</td>
<td>C1: 1:1 (before every attempt) C2: 2:1 (two models before every attempt) C3: 2:1 (two models, one before and one after every attempt)</td>
<td>c: consequating event (toy bear answered the question)</td>
</tr>
<tr>
<td>Zwitman &amp; Sonderman (1979)</td>
<td>Exact</td>
<td>Not specified</td>
<td>Ant</td>
<td>Model discontinued after 10 consecutive correct</td>
<td>1:1 (before every attempt)</td>
<td>c: “good” and tokens that could be exchanged for a gift</td>
</tr>
</tbody>
</table>

Note. Ant = antecedent; NA = not applicable; c = correct; i = incorrect; Cons = consequent; C1/C2/C3 = Condition 1/2/3.

Figure 5. Timing of the elicitation prompt relative to the child’s response.

a. Immediate imitation: Say “The boy is running”

![Immediate imitation diagram]

b. Delayed imitation: “The boy is running.” Now you say it.

![Delayed imitation diagram]
Overall, the studies showed an increase in performance following imitation training therapy and higher performance relative to a control group that did not receive any therapy. It is important to note that, in all but one study, targets were absent (i.e., produced with 0% accuracy) prior to therapy. However, the studies do not unequivocally link the increased performance to the imitation training therapy or consistently show generalization of production to conversational speech or untrained exemplars.

**Studies Comparing Imitation Training to Observational Modeling**

**Imitation Training With Antecedent Models**

Eight studies compared imitation training to observational modeling. Recall that, in imitation training, children both hear a model and attempt to produce the modeled utterance after the model, whereas in observational modeling, children hear a set of utterances with the target form (see Figure 3).

Three studies (Anderson, 2001; Connell, 1987; Connell & Stone, 1992) compared imitation training to observational modeling alone without any production attempts (illustrated in Figure 3b) for targets that were absent at baseline. In one of the studies, the child was instructed to just listen without saying anything during the observational modeling condition (Connell, 1987). In two other studies, the child produced a similar utterance that did not include the target form after hearing each model rather than just listening without saying anything (Anderson, 2001; Connell & Stone, 1992). To illustrate, in Connell and Stone (1992), the child first heard a model of the target form (“You tell...”).
Spinky to pick up the TVum.”, where “um” was an invented
morpheme that coded brokenness) and was then instructed
to give a command to the puppet Spinky. In the imitation
condition, the child was told “Say ‘Spinky, pick up the
TVum’” to elicit an imitation of the target form, whereas
in the observational modeling condition, the child was told
“Say ‘Spinky, pick it up’” so that the child would not say
the target form.

In both studies by Connell (1987; Connell & Stone,
1992), the rate of correct responding by children with DLD
was higher in the imitation training condition than in the
observational modeling condition, while the children with TL
either had a higher accuracy rate for observational modeling
(Connell, 1987) or showed no difference between conditions
(Connell & Stone, 1992). In contrast, Anderson (2001) re-
ported no difference in performance for children with DLD
between imitation training and observational modeling.

In three other studies (Courtright & Courtright,
1976, 1979; Haynes & Haynes, 1980), observational model-
ing was followed with elicited production (see Figure 3c).
The children were prompted to attempt production of the
target—either the same utterances that had been
modeled (Courtright & Courtright, 1976) or novel utter-
ances (Courtright & Courtright, 1979; Haynes & Haynes,
1980) after hearing the set of models. Courtright and
Courtright (1976) reported an initially higher rate of cor-
correct responding for imitation training followed by a plateau
in performance. The observational modeling condition plus
elicited production showed a lower initial rate of correct
responding followed by a steady increase in accuracy and
a higher final accuracy rate 1 week after the completion of
the therapy sessions. It is important to note that baseline
accuracy was higher for the imitation training group, rang-
ing from 10% to 65% compared to 0% usage for the obser-
vational modeling group. Courtright and Courtright (1979)
similarly reported a higher accuracy rate for the observational
modeling plus elicited production condition at the 1-week
follow-up but also found higher accuracy in each session
for targets that were absent at baseline for both therapy
conditions. In contrast, Haynes and Haynes (1980) reported
a higher rate of correct responding in all sessions for imita-
tion training but reported no difference at a 1-week follow-
up for targets that were absent or emerging (i.e., produced
with 10% accuracy) at baseline.

Overall, the studies contrasting imitation training
with observational modeling alone (i.e., which did not in-
clude production of the target form) suggest that providing
opportunities for children with DLD to attempt produc-
tion of the target form will result in better performance.
However, the studies do not unequivocally show imitation
training to be more effective than observational modeling
that includes production attempts. While some studies
suggested that imitation training (i.e., having imitative
attempts alternate with the model) might yield an initially
higher rate of correct productions for modeled utterances,
other studies showed that observational modeling with
elicited production (i.e., having the child first hear a set
of models before attempting nonimitative production of
the target) may be equally effective in increasing accuracy
for modeled utterances and might even result in higher accu-

rate over time for nonimitative production of untrained ex-
emplars. Note that, in four of these studies (Anderson, 2001;
Connell, 1987; Connell & Stone, 1992; Courtright &
Courtright, 1976), the target was an invented grammatical
form and we do not know whether this affected the outcome.

Imitation Training With Consequent Models

Only one study (Goldstein, 1984) compared imita-
tion training with consequent corrective models to observa-
tional modeling without production attempts. This study
included only children with TL and targeted an invented
grammatical form. In Goldstein (1984), two children only
learned the target forms in the imitation training condition.
The other four children showed more rapid acquisition of
the target forms in the imitation training condition than in
the observational modeling condition. This was consistent
with the studies of antecedent models; that is, performance
was better when children had to attempt production of the
target form. There were no studies that compared imitation
training with consequent models to modeling followed by
production attempts.

Studies Comparing Imitation Training
to Embedded Therapy

Imitation training provides concentrated opportunities
for the child to hear models of a target form and attempt
that form outside of a communicative interaction. Six stud-
ies (Camarata & Nelson, 1992; Camarata et al., 1994; Cole
& Dale, 1986; Friedman & Friedman, 1980; Hughes &
Carpenter, 1989; Nelson et al., 1996) compared imitation
training to embedded therapy approaches such as interac-
tive language development teaching (ILDT; Lee, 1974)
and conversational recast therapy (CRT; Baker & Nelson,
1984). Within these embedded therapy approaches, the
child hears and attempts the target form for communicative
purposes. However, the density of models and attempts is
lower than in imitation training.

ILDT is implemented during storytelling activities
and includes elicited imitation as one component of ther-
apy. Models for imitation are provided both prior to child
attempts and after child errors. Hughes and Carpenter
(1989) reported a slight advantage for imitation training
over ILDT—performance for three out of four participants
was higher in the imitation training condition. All four of
the targets in the imitation training condition reached crite-
rion compared to three of the targets in ILDT. However,
these authors did not evaluate whether the outcomes were
statistically different. In addition, only mean performance
was reported, and there was no indication about whether
outcome varied as a function of baseline levels, which var-
ied from 0% to 26% usage. There was little generalization
to conversation for either therapy.

In other studies, conversational recasting was imple-
mented during play activities. Children heard models only
after inaccurate attempts at the target form. These consequent
corrective models (termed “recasts”) added the missing target form to the child’s utterance, and there was no prompting for the child to imitate those models. In these studies, authors looked at both targeted productions during therapy activities (defined as imitative, elicited, or spontaneous productions of therapy targets during the therapy session in which that form was targeted) and nontargeted productions (defined as elicited or spontaneous nonimitative productions of the target form during a session in which that form was not being targeted). For targeted production during therapy sessions, there was either no difference between therapy conditions (Camarata & Nelson, 1992) or an advantage for imitation training (Camarata et al., 1994; Nelson et al., 1996). In these latter studies, children required fewer sessions and fewer therapy trials in the imitation training condition before producing the first targeted production of a grammatical form. In both Camarata and Nelson (1992) and Camarata et al. (1994), the first nontargeted production was achieved after fewer sessions and with fewer therapy trials for recasting than for imitation training. Furthermore, the time between achieving the first targeted production and the first nontargeted production was much shorter for conversational recasting than for imitation training (Camarata & Nelson, 1992), and some children did not achieve nontargeted production during the imitation training condition (Nelson et al., 1996). In contrast, Gillum et al. (2003) reported an advantage for conversational recasting for achieving a higher criterion on targeted productions (defined as five productions) as well as on achieving criterion for nontargeted production. Note that all of these studies included targets that were absent at baseline. Nelson et al. (1996) also included partially mastered targets, and these targets showed a larger increase in nontargeted production for the recasting condition.

Two studies looked at whether child differences in cognitive level (as measured by IQ) or language predicted which therapy would be more effective. The outcome measure for both studies was the Developmental Sentence Score (DSS; Lee, 1974) based on a conversational sample, and neither study specified the baseline level for therapy targets. In a study by Friedman and Friedman (1980) comparing imitation training to ILDT, children showed moderate gains in DSS after both therapy approaches, regardless of pretherapy language level. However, there was an interaction of therapy condition with IQ. Children with lower IQ showed higher gains for imitation training, while children with higher IQ made more gains in the ILDT condition. In contrast, Cole and Dale (1986) found no difference in DSS gain between imitation training and a conversationally based therapy approach that was similar to CRT, and neither IQ nor pretherapy language level predicted the amount of gain from either therapy.

Overall, the studies suggest that imitation training might be less effective for long-term learning than a conversationally based therapy approach for children with DLD. Children might start to produce the target form more quickly during imitation training. However, achieving a higher level of production in therapy and generalizing usage to conversational speech might be achieved more quickly with a conversationally based therapy approach. It is also, however, worth noting the short-term effect of imitation training on performance. Imitation training led to a rapid increase in production during therapy for target forms that were completely absent prior to therapy. It is also important to note that the advantage of conversationally based therapy might not hold for all grammatical forms. Camarata and Nelson (1992), for instance, noted that nontargeted production of gerunds was achieved more quickly in the imitation training condition.

Discussion

The reviewed studies show that imitation training increased production of grammatical forms in therapy. Specifically, children quickly learned to repeat the modeled utterances during therapy trials and generalized production to untrained exemplars. Importantly, this effect was seen for grammatical forms that were absent prior to therapy. The strongest evidence for this effect was from two Level 2 group studies that compared imitation training to CRT (Camarata et al., 1994; Nelson et al., 1996).

Imitation training involves two major components—hearing models of the target form and production attempts. The review suggests that production attempts are an important component of therapy (Anderson, 2001; Connell, 1987; Connell & Stone, 1992; Goldstein, 1984). However, it also suggests that production attempts need not be imitative (Camarata & Nelson, 1992; Camarata et al., 1994; Courtright & Courtright, 1976, 1979; Nelson et al., 1996).

We considered whether there was any best way to implement imitation training. We found only one study that compared different ways of implementing imitation training. This study showed increased efficacy for imitation training when the child heard more models (Valian & Casey, 2003). This apparent benefit of increasing the density of models might explain why observational modeling was at least as effective as if not more so than, imitation training when observational modeling included production attempts (Courtright & Courtright, 1976, 1979). However, the timing of those models might not be important. Specifically, it might not be necessary to provide models before production attempts, as was done in most of the studies. Consequent models have also been shown to be effective in increasing production of target forms (Goldstein, 1984).

Contingency of models might also not be crucial. Performance increased when antecedent models were provided before every attempt, regardless of accuracy, as was done in many of the studies. However, starting antecedent models only after errors and then eliminating the model after some number of correct was also effective (Hegde,
1980; Hegde & Gierut, 1979; Hegde et al., 1979; Hughes & Carpenter, 1989). Providing consequent models only after errors was also effective (Goldstein, 1984). The review showed that children may make fewer errors when models are given before every attempt. Yet, the success of contingent models suggests that errorless learning is not a necessary condition for children to improve in therapy.

There has been a long-standing concern about a potential lack of generalization to conversational speech for imitation training (e.g., Fey, 1986; Leonard, 1981; Paul & Norbury, 2012). Many of the studies did not evaluate this. Those that did confirm that this may be a limitation. Generalization to conversational speech may not occur without the implementation of procedures to achieve the transfer of usage to conversational speech contexts (Mulac & Tomlinson, 1977) or after considerably higher amounts of therapy than therapies that involved conversational contexts (Camarata & Nelson, 1992; Camarata et al., 1994). Thus, the benefit of imitation training—a rapid increase in the short term for usage in therapy—might be countered by a slower rate of change for achieving long-term goals for communicative use.

An important consideration is to match therapy approaches to the client as therapy approaches might not be equally effective for all children. Our review suggests that cognitive level might be important to consider. That is, imitation training might be particularly useful for children with lower cognitive levels (Friedman & Friedman, 1980). Our review also suggests that how children with TL respond to therapy might not be applicable to children with DLD. That is, imitation training may be useful to help children with DLD rapidly achieve production of target forms even if children with TL achieve production more readily with another therapy approach (Connell, 1987; Connell & Stone, 1992).

Therapy approaches might also not be equally effective for all grammatical forms. Of note is the finding by Camarata and Nelson (1992) that nontargeted production of gerunds, a low-frequency occurring form, was achieved faster with imitation training than with recast therapy. Imitation training might be particularly useful for such infrequently occurring forms. This may explain the higher accuracy rates for imitation training than for observational modeling in studies of invented morphemes (Connell, 1987; Connell & Stone, 1992).

Despite a long history of using imitation training for working on grammar goals, there has been little investigation into the best way for implementing this therapy approach to maximize efficacy. There are alternative ways of implementing imitation training, and we need studies to determine which implementation conditions might result not only in increased usage in therapy but also in increased generalization to untrained exemplars and to conversational usage. Several lines of inquiry would be valuable, including comparing partial to exact imitation, comparing immediate to delayed imitation, comparing antecedent to consequent models, and comparing contingent to noncontingent models (i.e., models given regardless of accuracy). We also need studies that manipulate the density of models and explore how this affects the relative efficacy of imitative versus nonimitative production responses.

There appears to be little recent interest in investigating the efficacy of imitation training. This may reflect an assumption that the efficacy of this therapy approach had already been established and that additional studies are not needed. Our review shows that this is not the case. The studies whose aim was to evaluate imitation training, all of which were from 1980 or earlier, did not unequivocally prove the efficacy of imitation training. The focus of more recent studies has been to show that other therapy approaches work better than imitation training. Interestingly, it is these studies—particularly the comparisons with CRT—that have been most illuminating about the impact of imitation on the production of grammatical forms.

In phonological disorders, researchers have found that increasing a child’s ability to imitate target sounds (termed “stimulability”) leads to greater gains in therapy (Miccio & Elbert, 1996; Rvachew et al., 1999). The same may be true for grammatical targets. That is, the rapid achievement of imitative production from imitation training might enhance the outcome of the other therapy approaches and lead to greater gains, particularly for forms that are absent at the start of therapy.

Therapy studies are typically set up in an either-or scenario to determine whether one therapy works better than another. Studies are also needed to investigate whether combining two therapy approaches would work better than using only one of the approaches. For instance, given the difference in immediate and long-term outcomes between imitation training and CRT, it may be that a therapy protocol that combines these two approaches would be more effective than either approach alone.

Table 10. Imitation training variables providing more and less support.

<table>
<thead>
<tr>
<th>Variables</th>
<th>More support</th>
<th>Less support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of match</td>
<td>Exact imitation</td>
<td>Partial imitation</td>
</tr>
<tr>
<td>Timing of prompt</td>
<td>Immediate imitation</td>
<td>Delayed imitation</td>
</tr>
<tr>
<td>Timing of model</td>
<td>Antecedent model</td>
<td>Consequent model</td>
</tr>
<tr>
<td>Density of models</td>
<td>Before every attempt</td>
<td>Reduced schedule of models (i.e., before every second, third, ... attempt)</td>
</tr>
<tr>
<td>Contingency of model</td>
<td>Provide model regardless of correctness</td>
<td>Provide model only after error</td>
</tr>
</tbody>
</table>
Concluding Thoughts

Applying evidence to clinical practice is a challenge. To be confident in the value of using a particular therapy approach, we need evidence that the therapy approach works. We have some evidence for this for imitation training, at least regarding production of target forms during therapy activities and for extension of production beyond the trained exemplars. We also want evidence that the therapy approach works better than other therapies and that, therefore, it is the best choice of approaches for working on a particular target with a particular client. The evidence regarding this is mixed for imitation training. Imitation training may be the fastest way to achieve initial production and a high rate of accurate production during imitative therapy trials. However, it may not be the most efficient approach for achieving the longer term aim of communicative use in conversation.

This does not mean that clinicians need to completely stop using imitation training. Rather, it suggests the need to modify how imitation training is implemented. Specifically, we may want to discontinue the practice of first training a form in imitation until a high criterion is reached before working on that same form within discourse. An alternative implementation would combine imitation training with conversationally based therapies, by first practicing the target during a brief imitation training activity and then immediately following that with an activity involving communicative use of the target form (Eisenberg, 2005, 2013, 2014). Such a combined approach has been successfully used in several studies (Fey et al., 1993; Owen Van Herne et al., 2018). This approach might be particularly useful for grammatical forms that occur less frequently in conversation.

Clinicians might also want to reconceptualize imitation training as involving a continuum of support rather than as an all-or-none phenomenon. In most studies, models for immediate exact imitation were provided before every attempt and then abruptly discontinued when the child achieved some criterion for imitative production on therapy trials. It may be that the lack of generalization may be due to this abrupt elimination of the model rather than to imitation training per se. An alternative would be to manipulate variables such as the degree of match between child response and model, timing of prompts and models, contingency of models, and density of models to reduce support as shown in Table 10. This would provide a hierarchy of scaffolding that could be individualized to client needs.

In conclusion, we believe that the use of imitation training, with suggested modifications, can continue to be a useful tool in an SLP’s toolbox. Further studies, such as those suggested above, will serve to refine the procedure and add to the evidence base to support our expert opinion.

References


