Responses made by late talkers and typically developing toddlers during speech assessments

ROSEMARY HODGES1, ELISE BAKER1, NATALIE MUNRO1 & KARLA K. MCGREGOR1,2

1Discipline of Speech Pathology, The University of Sydney, Sydney, Australia and 2Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA

Abstract

Purpose: Assessing toddlers’ speech is challenging. We explored responses made by late talkers and their typically developing peers in structured speech sampling contexts and determined if late talker subgroups could be identified.

Method: Twenty-six late talkers and 26 age-matched typically developing toddlers participated in an expressive phonology assessment and an elicited non-word imitation test. We quantified the breadth of toddler responses used in a subset of monosyllabic stimuli from the toddler phonology assessment and in the non-word imitation test. Correlational and cluster analyses were conducted.

Result: There were six response types: no response, protoword response, different verbal response, correct phoneme, common and uncommon phonological errors. Toddlers’ use of most of the response types correlated across the two sampling contexts. Use of the response types also correlated with several direct and parent-report assessments. There were significant group differences in the use of several response types in both sampling contexts. Five late talker subgroups were identified that presented with differing profiles of responses.

Conclusion: Toddlers respond in a variety of ways during structured speech sampling contexts. Responses made by late talkers offer insights about the nature of late talking and their heterogeneity. Implications for research and clinical management of late talkers are discussed.

Keywords: Responses, toddlers, late talkers, subgroups, assessment

Introduction

Five late talking toddlers, Aaron, Hayden, Eddie, Penny and Nathan were referred for a speech-language pathology assessment. During their assessment, the speech-language pathologist (SLP) showed each toddler a picture of a duck in a picture naming task and asked, “What’s this?” Aaron said “[dAK]”, Hayden said “[A]”, Eddie said “[b]”, Penny said “[o]” and Nathan said nothing at all. The SLP then said, “It’s a duck, you say duck.” Aaron said “[kAk]”, Hayden said “[A]”, Eddie said “[bai]”, Penny said “[a]” and Nathan said nothing. Following this, the SLP attempted to elicit some non-words via verbal imitation and similarly varied responses existed.

This scenario may sound familiar to those who have worked with late talkers (LTs)—toddlers with delayed expressive vocabulary acquisition despite otherwise typical development—and have attempted to elicit speech from them using structured assessments. There is an inherent assumption in such assessments that young children will either spontaneously name pictures or objects in response to simple questions, or imitate target words or non-words. However, toddlers do not always respond in ways that we, as SLPs or researchers, expect. When given an opportunity, they could make a range of different responses as seen across the five LTs introduced. Such diversity may be characteristic of the clinical population of LTs, given the recognised heterogeneity of their expressive language abilities (Desmarais et al., 2010). Diverse responses pose a challenge to clinicians and researchers who want to assess the emerging speech production/expressive phonology skills of LTs.

Recently researchers have developed structured assessments to measure expressive phonology abilities in children under the age of three [e.g. the Toddler Phonology Test (TPT); McIntosh & Dodd, 2008; Profiles of Early Expressive Phonological Skills (PEEPS); Williams & Stoel-Gammon, unpublished]. These address a long-standing clinical need given identified expressive phonology difficulties among LTs (e.g. Thal, Oroz, & McCaw, 1995;
Williams & Elbert, 2003). These assessments are picture or object naming tasks designed to elicit single word responses comprising a range of consonants, vowels and word-shapes. Unlike routine assessments typically used with pre-school and school-age children [e.g. Goldman-Fristoe Test of Articulation-2nd edition (GFTA-2); Goldman & Fristoe, 2000], toddler-specific assessments use stimuli that are likely to be within the vocabularies of 2-year-olds and, in the case of the PEEPS, use toys or naming of body parts and actions to elicit the stimuli rather than pictures. Toddlers’ responses are used to identify their consonant and syllable structure inventories and are usually analysed with respect to segment accuracy [e.g. proportion of consonants correct (PCC), proportion of phonemes correct (PPC), proportion of vowels correct (PVC)] and the presence of phonological error patterns (McIntosh & Dodd, 2008; Williams & Stoel-Gammon, unpublished).

Toddlers’ expressive phonology abilities may also be sampled with tests of elicited word/non-word imitation. Tests of elicited word/non-word imitation are of clinical value in identifying children with language impairments (e.g. Graf Estes, Evans, & ElseQuest, 2007). In toddlers, performance on tests of elicited word/non-word imitation positively correlates with vocabulary size (e.g. Chiat & Roy, 2007), and has diagnostic promise (Hodges, Munro, Baker, McGregor, & Heard, 2016; Stokes & Klee, 2009). The ability to imitate non-words involves a variety of processing abilities and possibly draws on long-term lexical knowledge. As such, it has been used as a measure of a variety of processes, including but not limited to, speech production/expressive phonology (see Coady & Evans, 2008 for review). There have been recent suggestions that in toddlers, performance on elicited non-word imitation tests may primarily reflect short-term processing abilities, in particular, speech production processes (Brandeker & Thordardottir, 2015; Hodges et al., 2016; Stokes, Moran, & George, 2013). Given that non-word imitation data are usually measured via speech accuracy such as PPC, this further indicates that such tests could be utilised as measures of expressive phonology. In the current study, we consider a monosyllabic non-word imitation test with accuracy measured by PPC as a type of toddler speech sampling context. It is, however, possible that because the stimuli are novel (not familiar), and because the elicitation context involves imitation (not naming), there may be differences (as well as similarities) in how toddlers respond compared to their responses during a toddler expressive phonology assessment. Thus, it may be worthwhile to include both types of speech sampling contexts—a routine, toddler phonology assessment, and an experimentally controlled non-word imitation test—for gaining greater insights into toddlers’ responses during assessments requiring speech production.

Although toddler expressive phonology assessments and elicited non-word imitation tests provide us with tools through which to examine toddlers’ speech production/expressive phonology, the full breadth of responses possible has yet to be considered. When toddlers have an opportunity to respond verbally, they might, make no response or they might say something other than the target word or non-word. Currently, these different types of responses would simply be scored as incorrect, and perhaps discarded as unusable data. However, these responses could be of value.

What types of responses could we expect from toddlers?
The types of responses that we might expect from toddlers were ascertained by looking at diverse literature on learning to talk in toddlerhood. We identified two broad possibilities: not responding or vocally responding.

Not responding. A toddler’s lack of response may be informative. No response could be a type of response that is worth recording and analysing. What do we know about not responding in toddlers?

- Not responding is a common challenge in toddler elicited word/non-word imitation tests, with substantial non-compliance rates reported (e.g. Dohmen, 2012; Stokes & Klee, 2009).
- In tests of elicited word/non-word imitation, the highest rates of non-compliance are seen in toddlers with the smallest vocabularies (Chiat & Roy, 2007; Dohmen, 2012; Stokes & Klee, 2009).
- Research from the perspective of temperament and social-pragmatic skills in toddlers also suggests that toddlers with small vocabularies may be less responsive (Bonifacio et al. 2007; Irwin, Carter, & Briggs-Gowan, 2002).
- Not responding could mean that a toddler’s opportunities for speech practise and feedback are forgone. As suggested by Pharr, Ratner, and Rescorla (2000) talkative toddlers are allowed the articulatory practise, auditory feedback and caregiver responses that may support further lexical development.

Responding. If toddlers make a vocal response, they could do one of two things: they could attempt to say the target word/non-word, or they could make a vocal response that is not an attempt at the target word/non-word. What do we about each of these possibilities in toddlers?

- If toddlers attempt the target word or non-word, they may make phonological errors. These could include errors that are developmentally common or less common. Examples are summarised in Table I.
- LT toddlers may make common and uncommon phonological errors more often than age-matched typically developing toddlers (TDs) as a result of their smaller consonant inventories and syllable structure inventories (e.g. Thal et al., 1995). Some researchers
have reported uncommon phonological errors in LTs (e.g. Tyler, 1992; Williams & Elbert, 2003).

- If toddlers make a response that is not an attempt at the target word/non-word, one possibility is what Stoel-Gammon (1989) identified as an “all-purpose protoword” (p. 217), a phonologically simple, default-like response such as [ga] or [gæ] used across all word attempts.

- It is unknown whether responses such as protowords or other types of vocalisations occur during structured speech sampling contexts. Moreover, although it may be expected, it is in fact, unknown whether such vocalisations occur with greater frequency among LTs compared to age-matched TDs.

Why consider the response types in LTs?

Recall our five toddlers, Aaron, Hayden, Eddie, Penny and Nathan, who were referred to an SLP as LTs. Despite this shared label, when provided with an opportunity to respond to a real word, first through naming and then elicited imitation, they provided a range of responses. The use of a simple, dichotomous scoring system for segment accuracy would conceal potentially significant differences between these five LTs. Analysing toddlers’ responses to include the full breadth possible may allow for differentiation between LTs, and importantly, insights into the nature of their difficulties with learning to talk, and perhaps, the most appropriate goals for intervention.

To summarise, various responses may be possible during structured speech sampling contexts but, at present, we lack systematic ways of identifying and utilising the full breadth of responses. This study presents a first exploration of responses made by toddlers in two sampling contexts. The broad and specific research aims were as follows:

**Research Aim 1: To explore the responses used by toddlers**

(a) Identify the breadth of responses used by toddlers in a subset of monosyllabic stimuli from a toddler expressive phonology assessment and in a monosyllabic non-word imitation test.

(b) Determine whether toddlers’ use of the identified responses across the two structured speech sampling contexts—the subset of stimuli from a toddler expressive phonology assessment and the non-word imitation test—are related.

(c) Determine whether the identified responses used in both speech sampling contexts are related to toddlers’ performance on direct and/or parent-report measures of language, cognition, expressive phonology, temperament and social-pragmatics.

**Research Aim 2: To explore the use of the identified responses in the two groups: LTs and TDs**

(a) Explore the proportional use of the identified responses made in the two sampling contexts in LTs and age-matched TDs.

(b) Compare the groups of LTs and age-matched TDs in their proportional use of the identified response types in both speech sampling contexts.

**Research Aim 3: To explore the value of the identified responses relative to LTs’ heterogeneity**

(a) Determine whether the proportional use of the identified response types by LTs across both speech sampling contexts could be mined in a cluster analysis to subgroup LTs, and if so, to describe the characteristics of these subgroups.

**Method**

**Participants**

Fifty-two toddlers aged 25–35 months (M = 29.79 months, SD = 3.53 months) were included. Half were LTs and half were age- and gender-matched TDs. Recruitment was via community advertisements/flyers within Sydney, Australia. Participants in both groups met the following inclusionary criteria: earned scores ≥16th percentile on the Receptive One-Word Picture Vocabulary Test-4th edition (ROWPVT-4; Martin & Brownwell, 2011); scored “not at risk” on The Bayley Scales of Infant and Toddler Development Screening Test (cognitive subtest) (Bayleys; Bayley, 2006); passed a newborn
hearing screener and no parent concerns regarding hearing reported; and came from monolingual English-speaking homes. Additionally, to be included in the LT group, the parent reported no concerns about general development apart from expressive language development and the toddler scored < 15th percentile on the MacArthur-Bates Communicative Development Inventories (MCDI; Fenson et al., 2006). To be in the TD group, the parent reported no concerns about general development including expressive language development, and the toddler scored ≥ 16th percentile on the MCDI.

Both the groups included 15 boys and 11 girls from a range of socio-economic backgrounds based on geographical socio-economic status as indicated by percentile rank on the Socio-Economic Indexes for Areas (SEIFA; Australian Bureau of Statistics, 2013). Groups did not differ significantly in SES percentile rank [LT group: M = 66.92 (28.5), range = 6–100, TD group: M = 68.88 (24), range = 8–100, t(50) = 0.268, p = 0.790] or age in months [LT group: M = 29.88 (3.5), range = 26–35, TD group: M = 29.69 (25–35), range = 25–35, z = −0.369, p = 0.712].

A number of direct and parent-report assessment measures were conducted. There were significant differences between the LT and TD groups on measures of expressive vocabulary as expected: MCDI number of words [t(37) = 11.856, p = 0.001] and the Expressive One-Word Picture Vocabulary Test-4th edition raw scores (EOWPVT-4; Martin & Brownwell, 2011) (z = −5.658, p = 0.001). In addition, significant group differences were evident in receptive vocabulary as measured by ROWPVT-4 raw scores (z = −4.052, p = 0.001), Bayleys cognitive screener raw scores [t(50) = 4.157, p = 0.001], total PPC for all items produced on the PEEPS (z = −5.621, p = 0.001), total consonant inventory as calculated based on all non-imitated productions in the PEEPS [t(50) = 6.019, p = 0.001] and, in line with previous research (e.g. Bonifacio et al., 2007), social-pragmatic skills as measured by parent report on the Social-Conversational Skills Rating Scale (SCSRS; Girolametto, 1997) (z = −4.549, p = 0.001).

However, in contrast with previous research (e.g. Irwin et al., 2002), there were no significant differences between groups in parent-reported child temperament as measured by the competence scale of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2006) (z = −1.932, p = 0.053). Descriptive statistics for participant characteristics and the direct and parent-report measures for each group can be seen in Table II.

Procedure and materials

The two speech sampling contexts that are the focus of this study were embedded in a larger assessment

Table II. Within group means, standard deviations and ranges for direct and parent-report assessments conducted.

| Group | Vocabulary assessments | Other assessments conducted |
|-------|------------------------|----------------------------|
|       | ROWPVT-4 (raw score)   | Total consonant inventory  |
| LTs   | 24.92 (7.54) 13–44    | 12.69 (5.63) 1–21          |
| TDs   | 35.69 (8.58) 22–54    | 18.58 (8.21) 14–22         |
|       | EOWPVT-4 (raw score)  |                           |
| LTs   | 5.70 (1.84) 5–10      | 3.41 (0.50) 2.2–4.3        |
| TDs   | 10.76 (9.28) 0–27     | 15.84 (18.96) 27–95       |
|       | PEEPS PPC             |                             |
| LTs   | 0.38 (0.23) 0.00–0.75 | 12.69 (5.63) 1–21          |
| TDs   | 0.81 (0.12) 0.00–0.95 | 20.85 (4.00) 11–28         |
|       | MCDI number of words  |                             |
| LTs   | 50.00 (49.49) 16–90   |                             |
| TDs   | 50.00 (49.49) 16–90   |                             |
|       | MCDI percentile       |                             |
| LTs   | 57.00 (5.63) 11.25–83 |                             |
| TDs   | 57.00 (5.63) 11.25–83 |                             |
|       | SCSRS                  |                             |
| LTs   | 5.18 (0.70) 3.00–7.00 |                             |
| TDs   | 5.18 (0.70) 3.00–7.00 |                             |
|       | BITSEA competence screener |                      |
| LTs   | 17.18 (2.22) 10–0    |                             |
| TDs   | 15.98 (2.17) 14–12   |                             |

Note. ROWPVT-4 = Receptive One-Word Picture Vocabulary Test-4th edition (Martin & Brownwell, 2011), MCDI = MacArthur-Bates Communicative Development Inventories (Fenson et al., 2006), EOWPVT-4 = Expressive One-Word Picture Vocabulary Test-4th edition (Martin & Brownwell, 2011), BITSEA = Brief Infant-Toddler Social and Emotional Assessment raw score / 22 (Briggs-Gowan & Carter, 2006), MCDI = MacArthur-Bates Communicative Development Inventories; EOWPVT-4 = Expressive One-Word Picture Vocabulary Test-4th edition; BITSEA = Brief Infant-Toddler Social and Emotional Assessment raw score / 22; SCSRS = Social-Conversational Skills Rating Scale mean of assertiveness/responsiveness raw scores / 5 (Girolametto, 1997).
Speech sampling context one: The toddler phonology assessment, PEEPS. The PEEPS was developed by Williams and Stoel-Gammon (unpublished) and is a play-based assessment designed to profile toddler expressive phonology skills using words familiar to toddlers. The word list for the PEEPS and permission to use the assessment were obtained from its developers (A. L. Williams, personal communication, July 2012). During the PEEPS, 60 real words were attempted to be elicited from each toddler. Productions of elicited stimuli were used to calculate their total PPC and consonant inventory size.

For the exploration of toddler responses reported in this study, a subset of eight monosyllabic CVC words was selected. These words were chosen as they were reported to be said by at least 86% of 25-month-olds based on the CLEX database (Jørgensen et al., 2010).

To elicit the stimuli on the PEEPS, the examiner typically asked the toddler to name objects/body parts/actions. If the toddler did not provide a response or provided an erred response spontaneously, they were usually given the opportunity to imitate (e.g. “it’s a duck, say duck”). However, in doing this, the examiner used clinical judgment to balance eliciting the toddlers’ best production with the level of toddler frustration, and the need to complete the entire assessment. Therefore, at times, a toddler may have been asked to imitate, without a prior spontaneous naming opportunity or, at times, a toddler may not have been asked to imitate after providing an erred spontaneous production. This flexibility reflects routine clinical elicitation procedures when conducting structured speech sampling with such young children. When examining the toddler responses to the eight selected monosyllabic CVC words, we analysed each toddler’s best production of these words.

Speech sampling context two: The non-word imitation test, the Monosyllable Imitation Test for Toddlers. The Monosyllable Imitation Test for Toddlers (MITT) was designed by the authors. Details about task development and design are reported elsewhere (Hodges et al., 2016). Briefly, the MITT was an animated computer game with two episodes; one presented at each home visit session. In each episode, four non-words paired with unusual toy referents were presented within the context of a pragmatically motivating story. In episode one, the story involved a penguin called Percy who needed to pack away his toys. The toddler was asked to help Percy pack away his toys in a box by saying the name of the toy. Each toy was seen one at a time while the toddler simultaneously heard a non-word three times via recorded voice. After the final exposure, the toddler was prompted to imitate the non-word. After the toddler had attempted to imitate, the toy referent moved into the box while a “yay” sound effect played. In the case where a toddler did not respond vocally, the examiner pointed to the toy referent on the screen and said: “you say…” If, after a further 10 s, the toddler did not respond, the examiner said the non-word, the referent moved into the box, and the game continued. Episode two made use of a similar script to that used in episode one except that the toddler needed to say each non-word to make unusual toy referents move into a shopping trolley so that Percy the Penguin could buy new toys. The episodes were counter-balanced across participants: half the participants saw episode one in the first home visit session while half saw episode two in the first home visit session.

Characteristics of the stimuli used: PEEPS and MITT. The syllabic structure of the subset of eight PEEPS stimuli and the eight MITT stimuli was the same—CVC. The real words from the PEEPS and non-words in the MITT contained earlier- and later- acquired consonants based on Chirlian and Sharpley (1982). Given that the eight PEEPS stimuli were a subset of real words from the PEEPS, the consonant complexity was not balanced between early- and later- acquired. Rather, a variety of consonants were present in both word-initial and word-final position. For the MITT, a variety of earlier- and later- consonants were present in word-initial position, but word-final was held constant as early-acquired. The consonants in word-initial and word-final position in the PEEPS and MITT stimuli are outlined in Table III.

Identifying response types and determining their proportional use. To identify the breadth of responses, the first two authors examined the data from the eight monosyllabic PEEPS stimuli and eight MITT stimuli. Figure 1 identifies the range of responses that were evident in the data and Table IV provides definitions of each of these identified response types. As anticipated, toddlers either provided no response, or they vocally responded. If they responded, they either: (i) attempted the target word/non-word and produced it correctly or with phonological error/s, or (ii) made a vocal response that was not an attempt at the target word/non-word. We further categorised the vocal responses that were not attempts at the target word/non-word into two types: a response that was reminiscent of Stoel-Gammon’s (1989) concept of a protoword, or as a “different verbal response” as detailed in Table IV.

To determine the toddlers’ proportional use of the response types, we devised a numerical system for coding responses. Each participant’s eight
monosyllabic PEEPS words and eight non-words were individually scored out of three—the number of phonemes in each stimulus item. If no response was provided, then all three points for that stimulus item were assigned to no response. If a response other than the target word/non-word was produced, then again all three points for that stimuli item were assigned to the relevant type of response—either “protoword” or “different verbal response”. If the target word/non-word was produced correctly, then all three points were assigned to “correct”. If the target word/non-word was produced with one or two accurate phonemes, then the relevant number of points was assigned to “correct phoneme”. If one or more phonological errors were made, each was examined and determined to be either common or uncommon, and points assigned accordingly. The response types were not mutually exclusive (e.g. on a single stimulus item, a toddler could produce one phoneme correctly, another using a common error and the third using an uncommon error). We then summed each participant’s response scores for all eight stimuli within each of the sampling contexts to derive a total raw score for each response type for the PEEPS subset of stimuli and the MITT. To obtain proportional data, we divided the total raw score for each response type by 24 (the total number of phonemes across all eight stimuli within each sampling context).

In making decisions about common versus uncommon phonological errors, we used the extant literature to guide the classification of consonantal errors. Additionally, a range of other consonantal substitution errors was evident in our sample that could not be explained by common phonological error substitutions or patterns. These errors were present either in just one toddler or at the most, four toddlers in the entire sample (i.e. <10% of total sample) and were therefore coded as uncommon. Examples of these uncommon errors can be seen in Table IV. For vowels, in the absence of literature to guide us in determining common vs. uncommon

**Table III. Consonants contained in the eight PEEPS and eight MITT stimuli in word-initial and word-final position.**

| Sampling context | Word-initial | Word-final |
|------------------|-------------|-----------|
| Early            | /n, h, d, d, b, k/ | /s, d, z/ |
| Late             | /t, p, k, g/ | /l, s, z/ |
| PEEPS*           | /n, d, p, n/ | /s, l, d, z/ |
| MITT             | /p, p, m, n, t, u/ | /p, p, m, n, t, u/ |

*The eight PEEPS words were a subset of CVC words from Williams and Stoel-Gammon’s (unpublished) toddler expressive phonology assessment, the PEEPS.
**Table IV. Identified response types and definitions.**

| Response type            | Definition                                                                 |
|--------------------------|---------------------------------------------------------------------------|
| No response              | No vocal response made.                                                   |
| Protoword response       | A vocal response but not an attempt at the target word/non-word form evident at least two different responses made by a toddler within one speech sampling task (i.e. within the PEEPS subset of stimuli or the MITT). In this study, a protoword is considered to be a default production used by an individual toddler comprising V, CV or VCV syllable shapes and nasal or plosive consonants (e.g. [na] or [ada]). |
| Different verbal response| A different verbal response that was not an attempt at the target word/non-word. In the MITT, it was either a toddler created non-word that followed the phonotactics of English such as [do] or [gi] or a real word that was somehow semantically related to the non-word or the context such as “teddy” or “box” or a demonstrative pronoun (e.g. “that one”, “there”). In the PEEPS, it was a real word or sound effect related to word or context, or a demonstrative pronoun. |
| Correct phoneme          | One or more phonemes in the stimulus item were produced correctly relative to the adult target phoneme. |
| Common phonological error| One or more phonemes in the stimulus item were produced using an error that is considered developmentally common in young children. These included: fronting of velars and fricatives, stopping of fricatives and affricates, deaffrication, gliding, voicing errors, assimilatory errors, final consonant deletion, reduplication and diminution. Vowel errors that were considered common and present in >10% of the sample were subtle length changes (e.g. [pæn] instead of /pæn/) or replacing the vowel with the neutral vowel (e.g. [d#p] instead of /d#p/). |
| Uncommon phonological error| One or more phonemes in the stimulus item were produced using an error that is considered developmentally uncommon in English (initial consonant deletion, backing) or any other substitution or word structure error present in <10% of our sample that could not be explained by common errors/processes. These included a wide range of consonantal substitution errors such as [fkn] instead of /kn/, [hn] instead of /vn/, and [mot] instead of /mot/. Vowel errors that represented an obvious substitution and that were present in <10% of the sample were also coded as uncommon. For example, [das] instead of /d̩p/ or [hnt] instead of /hnt/. Note: Three responses for two LTs were difficult to classify because they added word initial stressed syllables to their attempts (e.g. [ina] instead of /vna/) and as such all three phonemes comprising the word or non-word were classified uncommon in these cases. |

Errors, we used the following conservative method: subtle changes in vowel length (e.g. a/a and u/u) and substitutions of the neutral vowel (a) were present in >10% of our sample and so were coded as common. All other vowel errors were present in <10% of our sample and coded as uncommon. Examples are presented in Table IV.

**Reliability**

The first author re-transcribed 10% of the data from the PEEPS subset of stimuli and MITT and the second author independently transcribed 10% of the data. Intra- and inter-rater reliability for the PEEPS subset of stimuli was 0.91 and 0.89, respectively, and for the MITT was 0.92 and 0.90, respectively. The joint consensus of the first two authors was used to identify the breadth of response types and also employed to examine each toddler’s responses and assign points to the relevant response type/s.

**Result**

**Research Aim 1: To explore the responses used by toddlers**

The breadth of responses used by toddlers in the two speech sampling contexts (PEEPS subset of stimuli and the MITT) was identified as part of the method of this study, with six response types revealed (see Figure 1 and Table IV). Spearman bivariate correlations were run with all 52 participants to determine whether toddlers’ proportional use of the six identified response types (no response, protoword response, different verbal response, correct phoneme, common error and uncommon error) were associated with one another across the two sampling contexts. Significant positive correlations were found between responses types used in the PEEPS and the MITT for no response (r = 0.63, p < 0.001), protoword response (r = 0.58, p < 0.001) correct phoneme (r = 0.83, p < 0.001) and uncommon error (r = 0.39, p = 0.004) but not for different verbal response (r = 0.04, p = 0.80) or common error (r = 0.20, p = 0.15).

Spearman bivariate correlations were also run between all toddlers’ use of the six identified response types in the PEEPS and the MITT and their scores on the direct and parent-report assessments. Several of the identified response types used in both sampling contexts correlated with the direct/parent-report assessments, particularly those assessments involving speech production (e.g. the EOWPVT-4 raw score, the number of words on the MCDI, and total PPC and consonant inventory from the PEEPS). It was particularly interesting that accurately responding when producing monosyllabic words and non-words (i.e. using a high proportion of correct phonemes in the PEEPS or the MITT) was positively associated with direct/parent-report measures of expressive vocabulary and expressive phonology. On the other hand, responding inaccurately when producing monosyllabic words and non-words (i.e. using protoword responses or not responding in the PEEPS or the MITT) was negatively associated...
with other measures of expressive vocabulary and expressive phonology. Correlations between the responses types used in both sampling contexts and the direct/parent-report measures can be seen in Table V.

Research Aim 2: To explore the use of the identified responses in the two groups: LTs and TDs

Table VI shows the proportional use of each of the six identified response types in the PEEPS and the MITT for both groups. Interestingly, the TD group never used protoword responses in either sampling context, in the PEEPS they never used a different verbal response, and in the MITT, they always provided a response.

The LT and TD groups’ proportional use of the identified response types were then compared statistically using Mann–Whitney U-tests. For the PEEPS, the groups differed significantly in their use of protoword responses \( z = -3.461, p = 0.001 \), different verbal responses \( z = -4.268, \ p < 0.001 \), correct phonemes \( z = -5.398, \ p < 0.001 \) and uncommon errors \( z = -2.327, \ p = 0.02 \). The LTs used a significantly higher proportion than the TDs of no responses, protoword responses and uncommon errors, but a significantly lower proportion of correct phonemes.

Research Aim 3: To explore the value of the identified responses relative to LTs’ heterogeneity

An exploratory hierarchical cluster analysis was performed using the 26 LTs. We entered the LTs’ proportional use of the six identified response types in the PEEPS subset of stimuli and the MITT as clustering variables (i.e. 12 clustering variables in total). This revealed five clusters, or subgroups, that demonstrated differing response patterns. What follows is a description of each cluster regarding their proportional use of the response types in both sampling contexts. We also provide descriptive names for each cluster to identify what is most characteristic of each cluster of LTs.

Cluster A: “Good responders cluster”. This cluster contained six toddlers with a mean age of 32.70 months (5 males, 1 female) and represented the LTs who were most accurate in their responses across both sampling contexts. In the PEEPS, the
proportion of correct phonemes was 0.70 and in the MITT, it was 0.59. They also made some common errors (0.18 for PEEPS and 0.24 for MITT) and a small number of uncommon errors (0.05 for the PEEPS and 0.08 for the MITT).

Cluster B: “Adequate responders cluster”. This cluster had five toddlers with a mean age of 27.60 months (all female). It contained LTs who were relatively accurate in their responses during the PEEPS but demonstrated a range of different responses during the MITT. For the PEEPS, their proportion of correct phonemes was 0.75. In the MITT, their proportion of correct phonemes was lower at 0.25. During the MITT, they used a substantial proportion of different verbal responses (0.35), but this was not seen in the PEEPS (0.02). In addition to identifying the proportion of each type of response used by LTs in the five clusters, we explored the clusters in terms of their performance on a variety of the direct and parent-report assessment measures collected as part of the assessment protocol. These included the measures of cognition, receptive and expressive vocabulary, temperament, social-pragmatic and expressive phonology (including consonant inventory size and overall PPC in the PEEPS) (Note. The Supplement accompanying this article provides further details about the expressive phonological abilities of each LT toddler, in each cluster). Due to the small sample size and exploratory nature of the study, it was not possible to statistically compare clusters, and therefore, these measures were explored descriptively as seen in Table VII.

An exploration of other abilities in the LT clusters. No clear patterns appeared across the clusters for receptive vocabulary, cognition, temperament or social-pragmatic measures. However, what did emerge was a pattern in the mean scores of the five clusters in relation to assessments that involved speech production abilities (expressive vocabulary and expressive phonology). The “good responders cluster” and the “adequate responders cluster” had higher mean scores on production measures while the “varied responders cluster”, “protoword responders cluster” and the “no responders cluster” had lower mean scores on these measures.

Cluster E: “No responders cluster”. Five toddlers were included in this cluster with a mean age of 28.20 months (3 males, 2 females). The most notable characteristic of this cluster was the high proportion of no responses in the PEEPS (0.65). No responses were also relatively common in the MITT (0.27). In the MITT, they used a relatively high proportion of protoword responses (0.35), but this was not seen in the PEEPS (0.02).
Discussion

In this study, we explored the responses made by toddlers across two speech sampling contexts: a subset of monosyllabic stimuli from a toddler expressive phonology assessment and a monosyllabic non-word imitation test. A variety of response types were identified. Correlations showed interesting associations and dissociations in the use of response types across the sampling contexts, highlighting the value of including both contexts. For instance, no responses, protoword responses, correct phonemes and uncommon phonological errors seemed to be used by toddlers regardless of the sampling context, while different verbal responses and common phonological errors were not associated across the sampling contexts. More frequent use of different verbal responses in the MITT than the PEEPS is not surprising given the nature of the stimuli: non-words in the MITT versus highly familiar real words in the PEEPS. More frequent use of common phonological errors in the MITT may have reflected the toddlers’ lack of familiarity with the non-word stimuli. Previous research has identified that familiarity of words influence phonological accuracy (Keren-Portnoy, Vihman, DePaolis, Whitaker, & Williams, 2010). It was also revealed through correlational analyses that toddlers’ responses tended to be associated with their performance on direct and parent-report assessments in predictable ways. Such correlations supported the idea that toddlers’ responses were important and worthwhile to explore. Differences between the LTs and their age-matched TD peers in their proportional use of several of the identified response types were found. Furthermore, LTs’ responses were of value for identifying heterogeneous subgroups.

In the remainder of this discussion, we will reflect on the response types used and the insights they allow for, particularly what they might tell us about LTs underlying difficulties with developing an expressive lexicon. We will then consider the implications of this research for longitudinal studies of LTs, and clinical management of LTs.

Reflecting on the response types identified and their use by toddlers

Not-responding. Some toddlers did not respond to one or more words/non-words, and not responding occurred more frequently amongst LTs compared to TDs. These findings are consistent with other studies that have used structured speech sampling with toddler samples, including LTs (e.g. Chiat & Roy, 2007; Stokes & Klee, 2009). In fact, the TDs always provided a response in the MITT, and in the PEEPS, there were only four TDs who did not respond to one or more items. Thus, at least in these speech sampling contexts, and this sample of toddlers, not responding was relatively unusual for TDs, but more common in LTs.

The LTs who used the highest proportion of no responses (i.e. the “no responders cluster”) had some of the most limited expressive vocabulary and expressive phonology abilities in the sample of LTs (albeit slightly better than those in the “protoword responders cluster”). This finding is consistent with research suggesting that toddlers who are less voluble are likely to have the poor speech and language abilities (e.g. Pharr et al., 2000; Rescorla & Ratner, 1996). Their lack of responses, and manifestation as non-compliant LTs, may reflect underlying processing difficulties. Specifically, given their receptive vocabulary was within normal limits, it may be that these LTs struggle to form phonological and/or articulatory-phonetic representations for both familiar and new words that support word production. Stokes (2014) and Stokes, Moran, and George (2013) postulated that the nature of LTs’ difficulties with developing an expressive lexicon may be a problem with activating phonological representations for word production, and identified the articulatory and/or phonological networks involved in language processing as the areas in which breakdowns may occur.

For toddlers such as those in the “no responders cluster”, underlying processing difficulties may have subsequent consequences for their social interactions. Perhaps, such toddlers may have initially been more responsive, but many unsuccessful communication attempts resulted in decreases in their responsiveness over time. It is, however, also possible that the opposite could be true: that a lack of responsiveness was their primary difficulty initially, and this limited their opportunity for gaining experiences of processing and producing words. These are interesting theoretical ideas that could be addressed in future research with longitudinal study designs.

Responding vocally. Toddlers who most frequently attempted to produce the target word/non-word were TDs, LTs in the “good responders cluster”, and, to a lesser extent, LTs in the “adequate responders cluster”. Although TDs made some phonological errors when attempting the target word/non-word, phonological errors were more frequent amongst LTs. On inspection of the phonological errors used in the two sampling contexts, it was apparent that, for the PEEPS stimuli, LTs used significantly more common and uncommon errors compared to TDs, while for the MITT stimuli, LTs and TDs used similar numbers of common errors, but LTs used more uncommon errors. Thus, these findings are consistent with previous research that LTs frequently make phonological errors, including uncommon errors (e.g. Tyler, 1992; Williams & Elbert, 2003). However, uncommon errors were not completely absent from the TD group, suggesting that their existence alone does not necessarily signal
“atypicality” or pathology. This was also recognised by Stoel-Gammon and Dunn (1985).

It may be that the LTs in the “good responders cluster” and “adequate responders cluster” who most often attempted target word/non-word were more similar to, rather than different from, their TD peers: they were responding in ways that were mostly shared with TDs. They also showed relatively good expressive phonology abilities, and larger expressive vocabularies, compared to the other three clusters of LTs. Therefore, these toddlers’ difficulties with learning to talk appeared to be milder in nature. For these LTs, it is possible that, unlike the suggestions made by Stokes (2014) and Stokes, Moran, and George (2013), they do not have specific underlying difficulties with developing phonological and articulatory-phonetic representations of words that support production. Perhaps their underlying difficulties are related to mapping or creating links between different types of representations of words such as the link between a semantic and phonological representation. It may be that these LTs simply need greater exposure to words and their associated meanings to develop an expressive lexicon.

Some toddlers responded vocally, but it was a response akin to a protoword, in keeping with Stoel-Gammon (1989). It appeared to represent a means of being compliant in these structured contexts—the toddler knew they needed to say something—so he/she produced this easy-to-say response, regardless of the phonological form of the target word/non-word. Protoword responses were absent from the repertoire of responses used by age-matched TDs indicating that it may be a unique strategy employed by some LTs with highly limited expressive phonologies. For instance, protowords were relied upon almost entirely by the three toddlers in the “protoword responders cluster”. They used protowords when asked to produce familiar words and new words. For them, protowords may have become a way to manage the pragmatic aspects of tasks that require speech production, in the face of underlying difficulties with developing phonological and/or articulatory-phonetic representations of words required for production. Pragmatically, it represents a better alternative to not responding at all as frequently seen in the “no responders cluster” who also had restricted expressive phonologies. It was surprising then that differences between the “protoword responders cluster” and the “no responders cluster” were not apparent on the parent-report measure of social-pragmatics, although it could be that the measure was not sensitive enough, or differences might only become apparent with larger subgroups.

Another type of vocal response that did not represent an attempt at the target word/non-word was a different verbal response. Toddlers either produced their own non-words following the phonotactics of English, or real words that were related semantically. When responding to the non-words in the MITT, different verbal responses were present in both LTs and TDs. They may, therefore, represent a useful pragmatic strategy used by all toddlers when phonological and/or articulatory-phonetic representations for new word forms are too fragile to allow for word production. However, for the familiar words in the PEEPS, different verbal responses were absent from the response repertoire of TDs, suggesting that TDs had robust representations for these words including phonological and semantic information, likely acquired through many experiences with the word in everyday contexts. For the LTs, different verbal responses were used less frequently for the PEEPS words compared to the non-words but were still present in the repertoires of some. Therefore, some LTs struggle to build robust underlying representations even for highly familiar words, at least for naming purposes.

Implications for longitudinal research

Predictors that have clinical value in identifying which LTs are most at risk for persistent language and/or speech impairments continue to elude researchers (Dale & Hayiou-Thomas, 2013), but some abilities measured in toddlerhood have been identified as significant predictors of LTs’ long-term outcomes. In particular, receptive language has been implicated. For example, the receptive language abilities of LTs have been found to significantly predict their receptive and expressive language outcomes at ages 4–5 years (Chiat & Roy, 2008; Thal, Marchman, & Tomblin, 2013). However, given that such research has not excluded LTs with receptive language delays, it is not surprising that receptive language would predict outcomes. LTs that have receptive and expressive language delays likely represent their own subgroup of LTs with underlying difficulties that may be different from those with expressive delays only. In the search for predictors of outcomes for LTs with expressive only delays, it may be important to consider their production abilities. The nature of toddlers’ early productions of words/non-words could hold clues to understanding their long-term outcomes across the domains of both language and speech. Thus, longitudinal research following up subgroups of LTs identified based on a rich and detailed assessment of their production abilities would be welcomed, and could inform important clinical decisions about which LTs most require early intervention.

Implications for clinical management of LTs

For assessment, this study provides support for the recent call to assess expressive phonology in toddlers. As addressed by researchers such as McIntosh and Dodd (2008) and Williams and Stoel-Gammon (unpublished), structured assessments of expressive phonology for very young children are useful. Prior
to the development of toddler specific assessments, spontaneous speech collected during play was most often used as a means to assess expressive phonology in LTs (e.g. Paul & Jennings, 1992; Thal et al., 1995). However, unstructured assessments alone may fail to provide sufficient and consistent opportunities for toddlers to respond. In addition to assessing LTs’ responses during structured toddler expressive phonology assessments, it may be worthwhile to examine toddlers’ responses across multiple sampling contexts. In this study, we found that the toddlers’ proportional use of the identified response types differed somewhat across the two sampling contexts with certain types of responses more or less common when responding to the PEEPS stimuli versus the MITT stimuli. Thus, the use of non-word imitation tasks in complement with expressive phonology assessments may provide useful insights when assessing LTs. In fact, our five subgroups of LTs only became evident when their responses to both the PEEPS and MITT stimuli were considered in the cluster analysis. The responses from either sampling context in isolation did not result in differentiated clusters. While there are published examples of word/non-word imitation tests available for toddlers (e.g. The Pre-school Repetition Test; Chiat & Roy, 2007; Test of Early Nonword Repetition; Stokes & Klee, 2009), these tests were originally designed primarily to assess phonological short-term memory by increasing the length of the stimuli, not emerging segmental speech abilities. Nevertheless, such tests inherently involve speech production/expressive phonology. An alternative is to use monosyllabic non-words to limit the influence of phonological short-term memory processes, and the MITT offers one such tool.

A further assessment implication is a need to measure broadly and analyse the range of responses toddlers may make during structured speech sampling contexts. It is useful to examine segment accuracy and phonological errors in toddlers who are attempting to produce the words or non-words being elicited. However, this study demonstrated that many toddlers, particularly LTs, often respond in ways other than producing the word/non-word that we are attempting to elicit. We offer Figure 1 as a systematic way of examining the range of toddlers’ responses.

The cluster analysis reported in this study also has implications for intervention, and specifically for the development of treatment goals. For example, for LTs in the “no responders cluster”, perhaps intervention goals could initially focus on building responsiveness. For those in the “protoword responders cluster” and “varied responders cluster”, intervention goals may be focussed on developing their expressive phonologies by expanding their inventory of consonants and syllable structures while simultaneously working on carefully designed vocabulary targets that consider the toddlers’ phonological capabilities. Those LTs in the “good responders cluster” and “adequate responders cluster” may be appropriate candidates for more traditional LT goals that focus primarily on increasing expressive single word vocabulary and two-word combinations through changes to the language stimulation environment, increasing exposure and experience with words and word combinations (e.g. Girolametto, Pearce, & Weitzman, 1997). While we offer these initial thoughts on potential clinical implications for assessment and goal-setting, we recognise the exploratory nature of this study and the small sample of LTs included, and therefore further research is needed to validate the findings.

Future research directions

An interesting future direction would be to compare directly the responses that toddlers make when responding to familiar, real words versus non-words. We were unable to do this in this study due to the differences between the PEEPS and the MITT. Not only did these tasks differ regarding the lexicality of the stimuli, but there were also differences in the elicitation contexts meaning that any attempt to isolate and examine the influence of lexicality would have been difficult. Future research considering the influence of the phonological characteristics of the stimuli on responses made by toddlers would be valuable. Given that the characteristics of phonological neighbourhood density, phonotactic probability and segment complexity have been shown to influence toddlers’ accuracy when producing non-words/new words (e.g. Hodges et al., 2016; MacRoy-Higgins et al., 2013), it is likely that these would mediate the types of responses made by toddlers.

A further important future direction to help develop a greater understanding of the nature of LTs’ underlying difficulties with learning to talk would be to include younger, vocabulary-matched toddlers. Do younger, vocabulary-matched toddlers show similar response patterns to those used by LTs? However, answering this question may be difficult to do in reality, given less mature attention and cognition in younger, vocabulary-matched toddlers. As already identified, longitudinal research that follows subgroups of expressive-only LTs to determine which ones have persistent language and/or speech problems is needed.

Conclusion

This study revealed that toddlers, particularly LTs, do not always respond by attempting the target word or non-word during structured speech sampling contexts; rather they respond in a variety of ways. The responses used by LTs during assessments requiring speech production may hold value in understanding more about their underlying
difficulties with developing an expressive lexicon, and recognising that there are different manifestations of late talking.

Acknowledgements
The authors would like to thank the families and children that participated in this research. The first author would also like to acknowledge the support of the Nadia Verrall Memorial Research Grant from Speech Pathology Australia.

Declaration of interest
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

References
Australian Bureau of Statistics. (2013). Socio-economic indexes for areas [Online]. Retrieved from http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa.
Bayley, N. (2006). Bayley Scales of Infant and Toddler Development Screening Test (3rd ed.). San Antonio, TX: Pearson.
Bonifacio, S., Girolometto, L., Bulligan, M., Callegari, M., Vignola, S., & Zocconi, E. (2007). Assertive and responsive conversational skills of Italian-speaking late talkers. International Journal of Language and Communication Disorders, 42, 607–623. doi:10.1080/13682820601084386.
Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. American Journal of Speech-Language Pathology, 24, 126–138. doi:10.1045/2015_AJSLP-13-0106.
Briggs-Gowan, M.J., & Carter, A.S. (2006). BITSEA: Brief Infant-Toddler Social and Emotional Assessment. Bloomington, MN: Pearson.
Chiat, S., & Roy, P. (2007). The pre-school repetition test: An evaluation of performance in typically developing and clinically referred children. Journal of Speech, Language, and Hearing Research, 50, 429–443. doi:10.1044/1092-4388(2007/030).
Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. Journal of Child Psychology and Psychiatry, 49, 635–645. doi:10.1111/j.1469-7610.2008.01881.x.
Chirlian, N.S., & Sharpley, C.F. (1982). Children's articulation development: Some regional differences. Australian Journal of Human Communication Disorders, 10, 23–30. doi:10.3109/02198219810102.03.
Coady, J.A., & Evans, J.L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). International Journal of Language and Communication Disorders, 43, 1–40. doi:10.1080/1368282060116485.
Dale, P.S., & Hayiou-Thomas, M.E. (2013). Outcomes for late talkers: A twin study. In L.A. Rescrarla & P.S. Dale (Eds.), Late talkers: Language development, interventions, and outcomes. Baltimore, MD: Paul H Brookes Publishing Co.
Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2010). Three profiles of language abilities in toddlers with expressive vocabulary delay: Variations on a theme. Journal of Speech, Language, and Hearing Research, 53, 699–709. doi:10.1044/1092-4388(2009/07-0245).
Dohmen, A. (2012). An investigation of nonverbal imitation and language in children with specific language delay (Unpublished Thesis, City University, London).
Edwards, M.L., & Shriberg, L.D. (1983). Phonology: Applications in communicative disorders. San Diego, CA: College-Hill.
Fenson, L., Marchman, V.A., Thal, D., Dale, P.S., Reznick, J.S., & Bates, E. (2006). MacArthur-Bates communicative development inventories (2nd ed.). Baltimore, MD: Paul H. Brookes.
Girolometto, L. (1997). Development of a parent report measure for profiling the conversational skills of pre-school children. American Journal of Speech-Language Pathology, 6, 25–33. doi:10.1045/0149-0360.0604.25.
Girolometto, L., Pearce, P.S., & Weitzman, E. (1997). Effects of lexical intervention on the phonology of late talkers. Journal of Speech, Language, and Hearing Research, 40, 338–348. doi:10.1044/jslhr.4002.338.
Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe test of articulation 2 (GFTA-2). Circle Pines, MN: American Guidance Service.
Graf Estes, K., Evans, J.L., & Else-Quest, N.M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. Journal of Speech, Language, and Hearing Research, 50, 177–195. doi:10.1044/1092-4388(2007/015).
Grunwell, P. (1987). Clinical phonology. Baltimore, MD: Williams & Wilkins.
Hodges, R., Munro, N., Baker, E., McGregor, K., & Heard, R. (2016). The Monosyllable Imitation Test for Toddlers: influence of stimulus characteristics on imitation, compliance and diagnostic accuracy. International Journal of Language and Communication Disorders. Advance online publication. doi:10.1111/1460-6984.12240.
Irwin, J., Carter, A., & Briggs-Gowan, M. (2002). The social-emotional development of “late-talking” toddlers. Journal of American Academy of Child and Adolescent Psychiatry, 41, 1324–1332. doi:10.1097/00004583-200211000-00014.
Jørgensen, R.N., Dale, P.S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. Journal of Child Language, 37, 419–428. doi:10.1017/S030500090900954.5.
Keren-Portnoy, T., Vihman, M.M., DePaulis, R.A., Whitaker, C.J., & Williams, N.M. (2010). The role of vocal practice in constructing phonological working memory. Journal of Speech, Language, and Hearing Research, 53, 1280–1293. doi:10.1044/1092-4388(2009/09-0003).
MacRoy-Higgins, M., Schwartz, R.G., Shaffer, V.L., & Marton, K. (2013). Influence of phonotactic probability/neighbourhood density on lexical learning in late talkers. International Journal of Language and Communication Disorders, 48, 188–199. doi:10.1111/j.1460-6984.2012.00198.x.
Martin, N., & Brownwell, R. (2011). Receptive One-Word Picture Vocabulary Test-4 and Expressive One-Word Picture Vocabulary Test-4. Novato, CA: Academic Therapy Publications, Inc.
McIntosh, B., & Dodd, B. (2008). Two-year-olds’ phonological acquisition: normative data. International Journal of Speech-Language Pathology, 10, 460–469. doi:10.1080/1749500802149683.
McLeod, S., & Bleile, K. (2003). Neurological and developmental foundations of speech acquisition. Chicago, IL: American Speech-Language-Hearing Association Convention.
Pau, R., & Jennings, P. (1992). Phonological behavior in toddlers with slow expressive language development. Journal of Speech, Language, and Hearing Research, 35, 99–107. doi:10.1044/jsr.3501.99.
Pharr, A.B., Ratner, N.B., & Rescalona, L. (2000). Syllable structure development of toddlers with expressive specific language impairment. Applied Psycholinguistics, 21, 429–449.
Riscalona, L., & Ratner, N.B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). Journal of Speech, Language, and Hearing Research, 39, 153–165. doi:10.1044/jsr.3901.153.
Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. First Language, 9, 207–224. doi:10.1177/014272378900900607.
Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year-olds. *Topics in Language Disorders, 11*, 21–32.

Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Baltimore, MD: University Park Press.

Stokes, S.F. (2014). The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language, 41*, 634–657. doi:10.1017/S030500091300010X.

Stokes, S.F., & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language, and Hearing Research, 52*, 872–882. doi:10.1044/1092-4388(2009/08-0030).

Stokes, S.F., Moran, C., & George, A. (2013). Nonword repetition and vocabulary use in toddlers. *Topics in Language Disorders, 33*, 224–237. doi:10.1097/TLD.0b013e31829d058c.

Thal, D.J., Marchman, V.A., & Tomblin, B. (2013). Late-talking toddlers: Characterisation and prediction of continued delay. In L.A. Rescorla & P.S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169–201). Baltimore, MD: Paul H Brookes Publishing Co.

Thal, D.J., Oroz, M., & McCaw, V. (1995). Phonological and lexical development in normal and late-talking toddlers. *Applied Psycholinguistics, 16*, 407–424. doi:10.1017/S0142716400066017.

Tyler, A.A. (1992). Profiles of the relationship between phonology and language in late talkers. *Child Language Teaching and Therapy, 8*, 246–264. doi:10.1177/02664909200800302.

Williams, L., & Elbert, M. (2003). A prospective longitudinal study of phonological development in late talkers. *Language, Speech, and Hearing Services in Schools, 34*, 138–155. doi:10.1044/0161-1461(2003/012).