# Preschool children's consistency of word production

**Article** *in* Clinical Linguistics & Phonetics · February 2022

DOI: 10.1080/02699206.2022.2041099

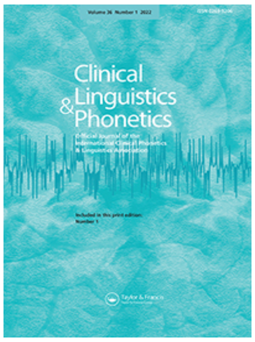


**6 authors**, including:

Alison Holm
University of Tasmania


Angela T Morgan
Murdoch Children's Research Institute


Barbara Dodd
The University of Queensland



Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

# Preschool children’s consistency of word production

Alison Holm, Olivia van Reyk, Sharon Crosbie, Simone De Bono, Angela Morgan & Barbara Dodd

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2022.2041099

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE



# Preschool children's consistency of word production

Alison Holm [a], Olivia van Reyk [b], Sharon Crosbie [c], Simone De Bono [b], Angela Morgan [b,d], and Barbara Dodd [b]

[a]School of Health Sciences, University of Tasmania, Launceston, Tasmania, Australia; [b]Speech and Language Group, Murdoch Children's Research Institute, Melbourne, Victoria, Australia; [c]School of Allied Health, Australian Catholic University, Brisbane, Queensland, Australia; [d]Audiology & Speech Pathology Department, University of Melbourne, Melbourne, Victoria, Australia

**ABSTRACT**

Consistency of word production contributes to carers' ability to understand children's speech. Reports of the proportion of words produced consistently by typically developing preschool children, however, vary widely from 17% to 87%. This paper examines the quantitative (consistency count) and qualitative (e.g. phonemic analysis) characteristics of word consistency in 96 children aged 36–60 months. Children named 15 pictures twice, in separate trials, in the same assessment session. The mean consistency of the production for the whole group was 82%. Older children were more consistent than younger children. Girls were more consistent than boys. Words produced correctly in one trial and in error in another may indicate resolving error patterns. Words produced in error in two different ways provided useful evidence about the nature of inconsistent word production in typically developing children. The clinical and theoretical implications are discussed.

**ARTICLE HISTORY**
Received 18 October 2021
Revised 14 January 2022
Accepted 1 February 2022

**KEYWORDS**
Consistency; phonology; word production; speech development; speech sound disorder

## Introduction

Typically, developing (TD) preschool children's ability to say words consistently (i.e. the same way each time) has been questioned. Reports of the proportion of words said consistently ranges from 17% (Jones, 2020) to 87% (Holm et al., 2007), although differing scoring methods may explain the different findings. Theoretical accounts attribute inconsistent word production to specific developmental abilities (e.g. motor-speech skills) or to a word's characteristics (e.g. word frequency, neighbourhood density; Jones, 2020; Vogel Sosa & Stoel-Gammon, 2006). Most critically, the clinical significance of inconsistent speech errors for the diagnosis of subgroups of speech sound disorder (SSD) is disputed (Sosa, 2015). This paper describes novel data from TD children's inconsistent word production to inform these issues.

Toddlers' early word production is highly 'variable' (Grunwell, 1982; Vogel Sosa & Stoel-Gammon, 2006). Before 24 months, words are acquired slowly, learned one by one, without awareness that words consist of a limited set of speech sounds that occur across words (Ferguson & Farwell, 1975). Ingram (1976) argued that once children have a vocabulary of 50 or more words, the nature of their productions changes and the rate of new word learning increases. Two-year-olds' words are produced more consistently, revealing emerging patterns

**CONTACT** Alison Holm alison.holm@utas.edu.au School of Health Sciences, University of Tasmania, Locked Bag 1322, Launceston, Tasmania 7250, Australia

of errors, shared across children learning the same language. The patterns reflect children's developing phonological knowledge, shown by constraints on word structure (e.g. final consonant deletion) and syllables (e.g. cluster reduction); and, substitutions, such as velar fronting (Smith, 1973). Consistent phonological pattern use by three-year-olds provides evidence about children's acquisition of language-specific phonological systems.

### *Consistency of word production in typically developing children*

In this paper, we consider a word 'consistent' when the same sequence of phonemes is produced each time it is uttered in the same context (e.g. picture naming). Phonetic variation in the production of a phoneme, captured by phonetic transcription, is not considered inconsistent. For example, *umbrella* [ɐmbelɐ], [ɐmbelə] is not phonologically inconsistent as the word final schwa is not a phonemic contrast and reflects normal variation in adult speech (Bürki, 2018). The speech signal that children hear and use as a model for speech production are high phonetically variable (Docherty & Mendoza-Denton, 2012). Toddlers are exposed to speakers of different ages, genders and language backgrounds and that influences their developing perception, internal representation, and production of sounds in words (Verspoor et al. (2021), resulting in some phonetic variation in children's speech.

Studies of word production consistency for TD children (aged ≥ 2;6) have most often used spontaneous picture naming tasks to elicit three separate tokens of between 20 and 30 words, obtained in the same assessment session. Measures include the proportion of words that were pronounced the same (e.g. all words correct; or all having the same error) or different (e.g. one correct and others in error; or two or more different errors). Studies using similar methodology are detailed in Table 1 (excluding reports of few words, only imitated words, or clusters).

**Table 1.** Percent consistency reported for typically developing children.

| Study | Age | N | Population | Mean % consistency: Overall and by approximate age group | | | | | | Gender effect |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Overall | ~ 2;6 | ~ 3;0 | ~ 3;6 | ~ 4;0 | > 4;0 | |
| Sosa (2015) | 2;6–3;11 | 32 | English (USA) | 32% | > 2;5 23% | > 3;0 32% | > 3;6 43% | | | Not tested |
| Macrae and Sosa (2015)[a] | 2;6–4;2 | 43 | English (USA) | 32% | | | | | | Not tested |
| Kehoe and Cretton (2021)[b] | 2;6–4;8 | 40 | French | 71% | | | | | | Girls > Boys |
| Jones (2020)[c] | 1;2 to 3;8 | 5 | English (USA) | 78% | | | | | | Not tested |
| Seunghee (2020) | 2;6–6;11 | 209 | Korean | 79% | > 2;5 44% | > 2;11 70% | | > 3;11 82% | > 4;11 88% | Not tested |
| Burt et al. (1999) | 3;10–4;10 | 57 | English (UK) | 82% | | | | > 3;9 77% | > 4;2 86% | Girls = Boys |
| Zarifian et al. (2020) | 3;0–6;0 | 317 | Persian | 84% | | > 2;11 81% | | > 3;11 87% | > 4;11 89% | Girls = Boys |
| Martikainen et al. (2019) | 3;0–6;11 | 80 | Finnish | 89% | | > 2;11 81% | | > 3;11 89% | > 4;11 96% | Not tested |
| Holm et al. (2007) | 3;0–6;11 | 409 | English (UK) | 92% | | > 2;11 87% | > 3;5 88% | > 3;11 93% | > 4;5 95% | Girls > Boys |

[a]Sosa's (2015) data combined with additional participants.
[b]50% of children were bilingual. Monolingual children's consistency: 74%; Bilingual children's consistency: 68%
[c]A longitudinal study; 1811 words analysed with two scoring methods over age range. Phonemic consistency: 78%; Phonetic consistency: 17%

A study of UK English-speaking four-year-old children ($n$ = 57) reported 82% word production consistency (Burt et al., 1999). Diagnostic Evaluation of Articulation and Phonology (DEAP, Dodd et al., 2002) normative data from 409 three- to seven-year-olds reported an average consistency of 92% on the Inconsistency Assessment. The DEAP data showed that three- and four-year-old children's consistency, in four 6-month age bands, rose from 87% to 95% (Holm et al., 2007). Similar levels of consistency are reported for same-aged TD children acquiring other languages (see, Table 1). Older children perform best and girls sometimes better than boys. In contrast, Sosa (2015) reported a consistency of 32% for 33 TD children aged 2;6–3;11 on the DEAP's inconsistency measure, a finding maintained when some additional children were included (Macrae & Sosa, 2015).

Jones (2020) investigated whether scoring of transcriptions might account for the discrepancy between Sosa's studies and other research. He analysed longitudinal spontaneous speech data from five children from the first words to 4;0 years. He first used a 'strict criteria' where phonetic transcriptions of a spoken word's productions were classified as correct only if they did not differ from the predetermined adult form, and consistent only if two error forms were phonetically identical. This approach yielded a mean consistency score of 17%. Perhaps this low score, compared to Sosa's (2015), p. 32%, reflects spontaneous speech as opposed to picture naming, the interval between data collection of first and second production (days, rather than the same session), different levels of transcription, and/or inclusion of data from children's first words.

Jones (2020, p. 128) second analysis of the same data considered 'production inconsistencies unlikely to be considered erroneous (e.g. [ælɪgeɪtɔɹ]/[æləgeɪtɔɹ] *alligator*; [mɒm]/[mʌm]) *mum*)' as consistent. This approach yielded a mean score of 78% consistency for his five participants, close to most studies of children aged ≥ 2;6 years (see, Table 1). Jones found that the difference between outcomes for the two scoring methods could account for the discrepancy in consistency scores reported by Sosa (2015) and other research. Few studies, however, provide transcribed examples of inconsistent production data that would allow evaluation of the nature of TD children's inconsistency.

### *Theoretical accounts of inconsistent word production in TD children*

One explanation for inconsistent errors focuses on the characteristics of words vulnerable to inconsistency. Bürki (2018) concluded that a word's neighbourhood density, lexical load, and frequency influences phonetic variation in adults. Similarly, Jones (2020) reported higher consistency for words with high neighbourhood density and lexical frequency, arguing that familiar words, from similar sound structure sets (neighbourhoods) are likely to have intact mental representations whose production is more practiced, 'requiring minimal articulatory and cognitive resources' (Jones, 2020, p. 145). Macrae and Sosa (2015) have also hypothesised that underspecified phonological representations, or difficulty accessing the mental lexicon, contributes to inconsistent speech errors.

Another explanation focuses on typically developing children's subtle motor-speech abilities (e.g. underlying phonetic planning and programming systems). While a word's mental representation may be intact, its production may still reflect immature motor programming and planning systems (Barbier et al., 2020). Sosa (2015, p. 32) suggests that rather than relying on phonetic transcriptions, 'refined acoustic and/or kinematic analysis methods may be needed . . . to assess accuracy and consistency' of word production.

A psycholinguistic perspective provides another account of inconsistent word production. Two separate speech output processing stages are identified in Levelt et al.'s (1999) model: a linguistic system incorporating concept, lexical choice, morphological and phonological planning; and a second stage, phonetic encoding of selected words as an articulatory motor program.

Evidence from speech errors, both normal adults' slips of the tongue and impaired speakers, suggest that phonological segments are assembled/planned before phonetic motor-speech gestures are programmed (Laganaro, 2019). In some cases of aphasia, people make inconsistent phonemic errors (paraphasias, e.g. *hedgehog* [dɪdʒɒg] [hɪdʒɒg], [ɛgɒg]) but have no motor speech-difficulties (Butterworth, 1992; Haley et al., 2013). Duffy et al. (2021) identified another group, without aphasia, who had primary progressive apraxia of speech characterised by articulatory distortions and poor prosody but no phonemic difficulties.

Similarly, there are two groups of children with SSD who make many inconsistent speech errors. Childhood apraxia of speech (CAS) is an impaired ability to program motor commands for articulating speech (Shriberg et al., 2012). Inconsistent phonological disorder (IPD) is another subgroup of SSD characterised by inconsistent errors. Around 10% of the children referred for speech difficulties have IPD, making inconsistent errors in the absence of any signs of CAS (Broomfield & Dodd, 2004). Most children with IPD benefit from 8 hours of therapy targeting consistency of production of a core vocabulary that generalises to untreated words (Crosbie et al., 2021). Dodd and McCormack (1995) attributed their difficulty to impaired phonological planning of phoneme sequences that make up words, in the absence of any speech-motor difficulty. In contrast, childhood apraxia of speech is attributed to motor planning/programming difficulties that manifest at the segmental level of speech (e.g. Maassen, 2002; Ozanne, 2005). Table 2 shows the symptom profiles of these two groups, including speech characteristics.

Sosa (2015) disputes the validity of inconsistent word production as a marker for SSD in the absence of symptoms of a motor planning disorder as found in CAS, claiming that inconsistency is typical of preschool children's word production. The differences in research

**Table 2.** Diagnostic characteristics of childhood apraxia of speech and inconsistent phonological disorder.

| Childhood Apraxia of Speech (CAS) | Reference | Inconsistent Phonological Disorder (IPD) | Reference |
|---|---|---|---|
| 3.4% of referrals with speech difficulties | a | 10% of referrals with speech difficulties | b |
| *Speech Characteristics for n = 8, mean age 6;1* | d | *Speech Characteristics for n = 31, mean age 4;3* | d |
| Consistency 36% | | Consistency 42% | |
| PPC 38% | | PPC 61% | |
| Mean number of distinctive features difference from target = 2.3 | | Mean number of distinctive features difference from target = 4.2 | |
| Imitation worse than spontaneous production | c | Imitation better than spontaneous production | e |
| Poor oromotor skills, articulatory groping | f | Oro-motor skills within normal limits. | e |
| Prosodic disturbance | a; c; f | Prosody good, normal affect | e |
| Dysfluent | | Fluent, | |
| Short utterance length | | Normal utterance length | |
| Syllable segregation | | No syllable segregation | |
| Slow speech rate | | Normal-rapid speech rate | |
| Restricted speech sound repertoire | | Age-appropriate speech sound repertoire | |
| Distortions of consonants and vowels | | Distortions rare | |
| Reading, spelling and writing difficulties | a | Good readers, poor spellers | e |
| Best therapy focuses on phonetic gesture | f | Best therapy focuses on phonological planning | e |

[a]ASHA (n.d.); [b] Broomfield and Dodd (2004), Delaney and Kent (2004), McCormack and Dodd (1998), and Crosbie et al. (2021), and McCabe et al. (2017).

findings concerning consistency/inconsistency in speech indicate that there is a need for valid data on consistency in typical speakers. The data would allow validation of subgroups of SSD. This study aims to quantitatively determine TD children's consistency of word production and to qualitatively (e.g. phonemic analysis) examine types of inconsistent errors.

### Research questions

Data from 96 children, aged 36–60 months, who produced 15 words twice, in two trials, addressed the following research questions.

(1) What proportion of the total number of words produced twice by all participants assessed was produced *the same* in both trials (either correctly or making the same error) *or differently* (either with one trial correct and one in error or making different errors in the two trials)?
(2) Is consistency in speech production associated with age?
(3) Is consistency in speech production associated with gender?
(4) When one word is correct and the other in error, what do qualitative phonemic analyses of inconsistent errors reveal about frequency and type of error patterns?
(5) What are the differences in phoneme production when a word is pronounced differently across trials?

The research reviewed lead to the hypotheses that typically developing children's word production would be predominantly consistent; older children would be more consistent than younger children; and girls would be more consistent than boys (see, Table 1). Qualitative phonemic analyses of types of inconsistent errors are important for comparison with inconsistent production of words by children with different types of SSD.

## Method

### Participants

Data for this study was drawn from a large normative study of speech development being conducted in Australia. Ethical approval was gained through the Royal Children's Hospital Human Research Ethics Committee (HREC 38243) and the Department of Education Victoria (Reference: 2018_003888). Fifteen of the 19 schools, kindergartens and childcare centres that were invited agreed to participate. Normative data collection was completed at nine sites. Parents/carers gave written consent for children's participation.

The 15 sites were stratified for socio-economic status (SES) using 2001 census data based on the Australian Bureau of Statistics' *Index of Relative Socio-economic Advantage and Disadvantage* (IRSAD ABS, 2016), classifying households into 5 classes from 1 (greatest relative disadvantage) to 5 (greatest advantage). Around a third of participants attended schools in the lower SES range (29%); middle range (33%); and upper range (38%).

Caregivers completed a health and developmental history during the consent process, documenting age, sex, Aboriginal or Torres Strait Islander status, primary language spoken at home and additional languages, history of speech pathology input, and health history and any developmental conditions (e.g. autism spectrum disorder and hearing loss). All 96

Table 3. Participant mean demographics by age group: age; gender; percent phonemes correct (PPC).

| Age group *months* | *n* | Mean age *months (SD)* | Girls *n* | Boys *n* | PPC *Mean (SD)* range | DEAP[a] PPC *Mean* |
|---|---|---|---|---|---|---|
| 36–42 | 24 | 40 (2.0) | 10 | 14 | 85 (14) 37–100 | 85–86 |
| 43–48 | 24 | 46 (1.7) | 12 | 12 | 90 (10) 66–100 | 88–91 |
| 49–54 | 24 | 52 (1.7) | 12 | 12 | 92 (10) 63–100 | 92–94 |
| 55–60 | 24 | 58 (1.4) | 12 | 12 | 93 (7) 71–100 | 96 |
| Total | 96 | | 46 | 50 | | |

[a]DEAP: *Diagnostic Evaluation of Articulation and Phonology* (Dodd et al., 2002) normative data, PPC range for a Standard Score of 10 for each age group.

children aged between 36 and 60 months with consent to participate were assessed. The only exclusion criterion was being non-verbal, as the assessment was for speech production. Of the 96 children assessed, 4 children had medical or developmental diagnostic information reported by their caregivers: 1 child had a repaired cleft lip; 1 had occasional dysfluency; 1 had a diagnosis of autism spectrum disorder; and 1 had a history of middle ear infections. An additional four children had been seen by a speech pathologist for assessment or intervention: two for unclear speech, one was late to talk and one did not specify the reason for consultation.

Information about accessing local speech pathology services was provided to parents when assessment raised concerns about a child's speech development. School closures due to the COVID-19 pandemic interrupted assessment, allowing analyses of data collected before April 2021.

Data reported here were from 96 participants aged between 36 and 60 months in four age groups with similar numbers of girls and boys (see, Table 3). The percent phonemes correct (PPC) scores, calculated from 30 words included in the consistency test, indicated that mean performance was within the normal range when compared with existing normal normative data (Dodd et al., 2002). The participants were selected in order of data entry, constrained by age and gender.

## Procedure

The data reported are from one assessment task, the Consistency Subtest, administered in a battery of speech assessment tasks. The children named 15 pictures, twice, in separate trials, to measure the consistency of word production. The 15 words selected were susceptible to speech errors in pilot trials. There were six one syllable words (cheese, sheep, witch, clouds, splash and frog); two words with two syllables (lunchbox, sprinkles); five three syllable words (screwdriver, elephant, octopus, kangaroo, strawberries (sometimes pronounced with two syllables [stɹo:bɹi:z]), and two words with four syllables (vacuum cleaner and helicopter). Their frequency of occurrence per million words ranged from 0 to 240 (Bååth, 2010) and neighbourhood density ranged from 0 to 28 (Bartolotti et al., 2012). A word structure score was derived by adding the number of syllables and two-element clusters plus 2 for every three-element cluster (e.g. *sprinkles* = 2 (syllables) + 1 (2-element cluster) + 2 (3-element cluster) = 5).

Assessments were done by four experienced speech pathologists who received training in the assessment protocol. Children were assessed individually in a quiet environment at their school, kindergarten or childcare centre. Assessment sessions for the complete battery of speech tasks were 30–60 minutes long, including breaks. Children wore lapel microphones.

Responses were transcribed 'live' using the International Phonetic Alphabet and compared to Standard Australian English vowels using the Harrington et al. (1997; HCE) system. Phonetic distortions were noted.

The children were shown coloured pictures of 15 words, one-by-one, and asked what it was, to elicit the target word spontaneously. If no response or the wrong word was given, a cueing hierarchy was implemented: semantic cue (e.g. target word *clouds*: 'up in the sky you see … '); forced choice (e.g. 'are they clouds or a plane?') and imitation (e.g. 'say clouds'). The second production of the 15 words was administered using the same procedure. Encouragement was provided during the assessment, but no feedback was given.

### *Reliability*

Children's naming was audio and video recorded for reliability checking. Data from ten participants (five under 43 months, four children approaching, or recently turned 48 months with one child over 54 months) were randomly selected. The videotapes were transcribed independently by two speech pathologists. Mean point-to-point reliability was 95.05% (SD 3.7, confidence interval ± 2.3%). Common conflicts affected cluster transcription (40%), such as schwa insertion (e.g. dwɑev vs. dəwɑev) or deletion of one member of the cluster (e.g. stwɔ vs. swɔ), and glide identification (26%) where /ɹ, w, l, j/ were sometimes perceived as another glide.

### *Analysis*

Each child's two productions of the 15 words were compared to quantify the number of pairs said the same (i.e. consistently) and differently (i.e. inconsistently). These two categories were then divided further into *Same*: both correct; both with same error/s; and *Different*: one correct and one with error/s; two words with different error/s. Qualitative analysis of phonemic errors described all changes in differently pronounced words. Phonological error pattern use was compared to normative data from the Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2002). The error patterns identified were classified as age appropriate (used by ≥10% of the children in the same 6 months age group); delayed (used by ≥10% of the children in a younger age group); or atypical (not used by ≥10% of the children in any age group in the normative data for children aged 2–7 years). Additional analyses examined how a words' characteristics (frequency, neighbourhood density and word structure) affected consistency of production.

## Results

### *Overview*

An overview of the consistency of speech production data from 96 children aged between 36 and 60 months is shown in Table 4. The 96 children produced 15 word pairs each resulting in a total dataset of 1440 pairs. The first research question concerned the number of words said the same in both trials (correctly or making the same error) in contrast to the number of words said differently (correctly in one trial and in error in the other; or, incorrectly in both trials, making different errors). Children said words the same way on 82% of the trials

**Table 4.** Same and different productions for 15 words said twice for 96 children aged 36–60 months.

| | Example | Mean (SD) | Range | % (of 15) |
|---|---|---|---|---|
| Combined Same | | 12.0 (2.9) | | 81.7% |
| Same: *Both correct* | *cheese* [tʃi:z] [tʃi:z] | 8.5 (4.9) | 0–15 | 58.8% |
| Same: *Same error/s* | *octopus* [æpəpʊs] [æpəpʊs] | 3.5 (3.6) | 0–15 | 22.9% |
| Combined Different | | 2.8 (2.6) | | 18.3% |
| Different: *1 correct, 1 with error/s* | *sheep* [ʃi:p] [si:p] | 1.3 (1.4) | 0–5 | 8.7% |
| Different: *With different error/s* | *strawberries* [tɹo:bi:bewi:z] [bwo:wi:z] | 1.5 (2.0) | 0–9 | 9.6% |

**Table 5.** 'Combined Same' frequency scores for 96 children aged 36–60 months across 15 words.

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of words both correct or both same error | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Total = 15 words |
| Number of children with this score | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 4 | 2 | 6 | 7 | 13 | 10 | 10 | 14 | 25 | Total = 96 children |

(59% correct; 23% same error). Children said words differently on 18% of the trials (9% one correct and one in error; 9% different errors). A paired t-test comparing the same and different scores indicated children's speech production of the same words was highly consistent ($t(95) = 16.599$, $p < .001$). The frequency distribution of scores for the same responses shows predominantly high scores (see, Table 5). Twenty-five children (26%) said all 15 words the same way on both trials and 75% of the children said 11–15 words the same.

## *Effect of participant characteristics on consistency of word production*

A two-factor analysis of the variance explored the effect of age (36–48 vs. 49–60 month olds) and gender (girls vs. boys) on the consistency of word production, using the combined same scores (both correct and same error). Table 6 presents data based on age and gender. The mean number of the 15 words pronounced consistently in two trials by the older group was significantly higher than for the younger group ($F(1, 92) = 8.033$, $p < .01$). Girls' performance was more consistent than boys' ($F(1, 92) = 6.462$, $p < .025$). The interaction term was non-significant ($F < 1$) indicating that girls' consistency was better across the age range.

Data from the eight children with medical or developmental case history information was inspected for individual differences in consistency from the group: six children had combined the same scores similar to the whole group mean score of 12 (their scores were 11, 11, 12, 13, 13, 14). Children with a diagnosis of autism spectrum disorder had a combined same score of 7 and the child who was a late talker had a combined same score of 5. The aim of the study was to look at consistency in a normative sample of children, and therefore data from all children were retained in the analyses.

**Table 6.** Consistency of 15 words produced twice by age and gender.

| | 36–48 months *Mean (SD)* | | 49–60 months *Mean (SD)* | |
|---|---|---|---|---|
| | Girls | Boys | Girls | Boys |
| Combined Same | 12.3 (2.4) | 10.4 (3.1) | 13.3 (2.2) | 12.4 (2.9) |
| Same: *Both correct* | 9.1 (5.1) | 6.7 (4.2) | 11 (3.6) | 7.9 (5.3) |
| Same: *Same error/s* | 3.2 (4.1) | 4.0 (2.9) | 2.2 (2.8) | 4.5 (4.1) |
| Combined Different | 2.6 (2.4) | 4.1 (2.6) | 1.8 (2.2) | 2.3 (2.9) |
| Different: *1 correct, 1 with error/s* | 1.2 (1.3) | 1.8 (1.5) | 1.1 (1.5) | 1.2 (1.4) |
| Different: *With different error/s* | 1.5 (2) | 2.3 (2.1) | 0.6 (1.2) | 0.9 (1.6) |

Analyses of differing productions are more complex. The mean for different errors was low, with relatively high standard deviations reflecting the data's variability (see, Table 6). Consequently, the characteristics of inconsistent speech errors for words' differing productions require qualitative phonemic analyses, using non-parametric statistics where appropriate.

### *Analysis of errors when one word was correct and the other in error*

The frequency of types of errors for the younger and older groups is shown in Table 7. There were 130 examples of word pairs produced correctly on one trial and in error on the other although 35% of the younger and 42% of the older children made none. There was no difference between the first and second production for when a word was correct: Trial 1, 51% correct; Trial 2, 49% correct. There were a total of 151 phoneme errors from the 130 words: 81 from 36–48 month olds and 70 from the 49–60 month olds. The proportion of words containing only one error was 91% for younger and 75% for older children (e.g. *sheep* [si:p]). Multiple errors in one word were; then, less common (e.g. *sprinkles* [kɹɪmkəlz]; *kangaroo* [tændəɹʉ:]).

The phonological patterns most frequently used by children were age-appropriate according to normative data (Dodd et al., 2002). These patterns included gliding, cluster reduction of three-element clusters and deaffrication (until 5;0 years); fronting of fricatives and velars and weak syllable deletion (until 4;0); and stopping (until 3;6). The high number of assimilation errors, usually suppressed by 3;0 years, was unexpected. These error patterns occurring less often (final consonant deletion, vowel and voicing errors) are uncommon after 3;0. Non-developmental errors (additions, backing, initial consonant deletion and affrication) were least common. In summary, 70% of the errors were age appropriate; 19% were delayed errors; and, 11% were atypical. When a word was correct on one trial and in error on the other, errors are most often developmental and usually age-appropriate.

**Table 7.** Words produced differently – correct in one but in error in the other: frequency of phonological patterns.

| Phonological patterns: *examples* | 36–48 months | 49–60 months |
|---|---|---|
| Gliding: *frog* [fwɔg]; [flɔg] | 25 [a] | 18 [a] |
| Cluster reduction: *lunchbox* [l nbɔks]; *frog* [fɔg] | 21 [a] | 16 [b] |
| Assimilation: *octopus* [ɔpəpʊs] | 8 [c] | 9 [c] |
| Deaffrication: *cheese* [ʃi:z] | 4 [a] | 5 [a] |
| Fronting fricatives: *sheep* [si:p] | 4 [a] | 6 [a] |
| Fronting velars: *kangaroo* [tændəɹʉ:] | 3 [a] | 2 [c] |
| Weak syllable deletion: *helicopter* [helkɔptə] | 3 [a] | 1 [c] |
| Stopping: *vacuum cleaner* [bækjʉ:mkli:nə] | 1 [a] | 1 [c] |
| Final consonant deletion: *cheese* [tʃi:] | 2 [c] | 4 [c] |
| Voicing: *frog* [bɔg] (example co-occurs with stopping) | 1 [c] | 0 |
| Syllable initial consonant deletion: *helicopter* [eli:kɔptə] | 3 [d] | 1 [d] |
| Addition: *elephant* [eləplent] | 2 [d] | 2 [d] |
| Vowel errors: *helicopter* [hɔli:kɔptə] | 1 [d] | 3 [d] |
| Backing: *strawberries* [ko:bi:z] | 2 [d] | 1 [d] |
| Af/frication: *sheep* [tʃi:p] | 1 [d] | 1 [d] |
| *Total errors* | 81 | 70 |

[a]Age-appropriate.
[b]Triclusters only.
[c]Delayed.
[d]Error patterns occurring in < 10% children in normative data from 716 children aged 2;0–6;11 (Dodd et al., 2002).

## Analysis of words pronounced differently, both in error

There were only 143 examples of a word being pronounced differently, with both in error: 94 made by the younger and 49 by the older group. No such errors were made by 51% of the participants. The Mann-Whitney U Test found that 49–60 month-olds ($n$ = 48 per age group) made fewer different errors than the 36–48 month-olds ($z = 2.3558$, $p < .025$).

The number of differences per word pair by phoneme count ranged from 1 to 5, with a mean of 1.7 (*SD* 1.1) for the younger and 1.8 (*SD* 1.1) for the older group. Most differences between words in a pair involved only one target consonant (e.g. *frog* [fɔg], [fwɔg]; *kangaroo* [tæŋgəwʉ:], [kæŋgəwʉ:]). Words eliciting two and three differences often reflected difficulties with three member clusters (e.g. *screwdriver* [ʃʉ:dɹɑevə], [dʒʉ: dʒɑevə]; *sprinkles* [fɪŋkəlz], [swɪŋkəlz]). Only 10 word pairs involved more than three phoneme differences. For example, *helicopter* [hɔkdətelə] and [hepətɔktə] scored five because only four of the nine sounds (/h, ə, t, ə/), were shared in sequence.

Table 8 shows the number of children making different errors when saying each stimulus word sampled twice, plus that word's frequency, neighbourhood density and phonological structure. Spearman's correlation coefficients revealed a strong positive correlation between the number of children making inconsistent errors on a word and its word structure ($r_s$ = 0.8348, $df$ = 2, $p < .001$); a borderline negative correlation between inconsistent error occurrence and neighbourhood density ($r_s$ = −0.5884, $df$ = 2, $p$ = .05); and no correlation with word frequency ($r_s$ = −0.352, NS).

Another analysis inspected the pairs of words with different errors to identify error types (see Table A1). Each word pair exhibiting different errors was classified into one of the three categories according to error characteristics. Out of the 143 examples of word pairs with

**Table 8.** Effect of word characteristics on inconsistent production: frequency, neighbourhood density and syllable structure.

Different: *With different error/s* (143 examples of a word being pronounced differently, with both in error)

| Word | Syllable and consonant/ vowel structure | Number of children with two different errors ($n$ = 96) | Frequency/ million words[a] | Neighbourhood density[b] | Word structure[c] |
|---|---|---|---|---|---|
| cheese | CVC | 1 | 240 | 25 | 1 |
| sheep | CVC | 1 | 59 | 21 | 1 |
| witch | CVC | 2 | 85 | 28 | 1 |
| clouds | CCVCC | 4 | 21 | 21 | 3 |
| kangaroo | CVC.CV.CV | 4 | 52 | 1 | 3 |
| octopus | VC.CV.CVC | 6 | 0 | 0 | 3 |
| elephant | V.CV.CVCC | 6 | 160 | 3 | 4 |
| lunchbox | CVCC.CVCC | 7 | 9 | 0 | 4 |
| splash | CCCVC | 9 | 44 | 3 | 3 |
| frog | CCVC | 10 | 87 | 5 | 2 |
| vacuum cleaner | CV.CCVC.CCV.CV | 13 | 7 | 1 | 6 |
| helicopter | CV.CV.CVC.CV | 17 | 36 | 1 | 4 |
| sprinkles | CCCVC.CVCC | 20 | 8 | 4 | 5 |
| screwdriver | CCCV.CCV.CV | 20 | 27 | 0 | 6 |
| strawberries or [stɹo:bɹi:z] | CCCV.CV.CVC or CCCV.CCVC | 23 | 68 | 0 | 5 |

Sampled vowels: ɪ i: ʊ ɔ æ ə e o: ʉ: ɑe; Consonants: p b t d k g m n ŋ f v s z ʃ tʃ ɹ l j w h
Clusters: syllable initial: br dr fr kl kj spl spr skr str; syllable final: dz ks lz ntʃ nt

[a]Bååth (2010); [b] Bartolotti et al. (2012); [c] Score derived by adding the number of syllables and two-element clusters plus 2 for every three-element cluster (e.g. *sprinkles* = 2 (syllables) + 1 (2-element cluster) + 2 (3-element cluster) = 5).

different errors, 80 of the word pair differences reflected alternative developmental phonological patterns: (e.g. gliding /ɹ/ [skwʉ:dɹɑevə] and cluster reduction [skʉ:dɹɑevə]; devoicing of post-vocalic /g/ [fɹɔk] and cluster reduction [fɔg]). Shared errors were not analysed (e.g. in *lunchbox* [lɐmpɔk] [lɐmpɔks] only cluster reduction counted). For word pairs with different errors, more than half were developmental (younger group 56%; older group 55%).

An obvious category to examine were pairs with atypical errors. There were 32 word pairs (22%) where inconsistency involved an atypical (non-developmental) error. The most common trend was af/frication of clusters (e.g. *strawberries* [ʃo:bewi:z] [sto:bewi:z]; [so:bi:z] [tʃoˑbɹi:z]; *sprinkles* [fɪŋkəlz] [swɪŋkəlz]). Other patterns included vowel errors, backing of plosives and initial consonant deletion.

The remaining 31 errors (22%) are shown in the Appendix under the heading of 'differences in word pairs due to syllable and word structure errors' including metathesis [hɔktətelə] [hetətɔktə]); partial syllable replication [kwæpʉ:kwi:nə] [kwæŋkʉ:nkwi:mə]; syllable and phoneme deletions and additions. Except for one, all the word pairs in this category were multi-syllabic. While interference from another word was evident in one word pair (*octopus* [ɔptəpʊs] [heliˑpʊts]), all resembled the target word to some extent. For younger children, 23% of the different errors were in this category compared to 18% for the older children.

## Discussion

The consistency of typically developing preschool children's word production is debated, poorly understood theoretically, and its clinical significance disputed. This report analysed 96 36–60 month old participants' productions of 15 words, in two separate trials. Overall, word consistency was 82%. Girls' single-word production was more consistent than boys, and younger children were less consistent than older children.

Qualitative analyses revealed that when one trial was accurate and the other in error that error was usually an age-appropriate developmental error. When both productions were in error, in different ways, 56% were both developmental errors (cheese [ti:z, ʃi:z]); 22% included a non-developmental error (*sprinkles* [kwɪŋkəlz, f̠wɪŋkəlz]); and 22% showed impaired sequencing of phonemes that nevertheless resembled the target (*strawberries* [tɹo:bi:bewi:z, bwo:wi:z]). Findings are now being considered in relation to the research questions.

### *What proportion of words sampled twice are consistent?*

The proportion of words pronounced the same in both trials was 82% (57% same correct, 25% same error), while words that were produced differently (one correct and one in error or with different errors) accounted for 9% each. Despite only sampling 15 mainly multisyllabic words twice, these findings reflect those of most previous studies (see, Table 1). The exception is Sosa (2015) who reported only 32% consistency. Jones (2020) evaluated this discrepancy, suggesting that Sosa's (2015) findings reflect different levels of transcription, and/or scoring criteria. Phonemic analyses consider allophones that vary phonetically (e.g. /p/ in pin vs spin), as the same phoneme that has the function of distinguishing word meaning (for discussion see, Ball, 2016). Sosa's (2015) call for instrumental analysis of children's speech to identify 'variability' infers a phonetic analysis, raising the issue of whether differences in a word's production that may not be perceived, and do not distinguish meaning, are clinically significant.

Bürki (2018) supports the use of instrumental measures to describe variation in the phonetic encoding of the motor program for articulation, the final speech output processing stage of Levelt et al.'s (1999) model. Before articulation is planned, however, a linguistic processing stage selects the word to be produced, assigns any necessary morphological marker, and plans the sequence of phonemes for production. This ability to assemble phonology is critical for the development of consistent production of words. The normative data presented here indicates that TD children are usually consistent in the way, they say words, maximising their chances of being understood, even when a word is in error.

## *Is consistency in speech production associated with age and gender?*

As hypothesised, older children produced words more consistently than younger ones, and girls were more consistent than boys across the age range (see, Table 6). Although statistically significant, differences were small. Older children produced a mean of around two words more consistently than younger children, and girls outperformed boys by one. Preschool children's communication abilities improve with age, as shown by standardised assessments. Research indicates that girls aged 3;0 to 5;0 years do slightly better than boys on many language measures, perhaps due to socialisation (Bornstein et al., 2004). However, girls outperformed boys across 10 diverse languages and cultural contexts where overall language abilities differed, suggesting their advantage is neither language specific nor culturally determined (Eriksson et al., 2012). Consistency of word production, then, likely reflects neuro-developmental change as well as preschool language experience. A ceiling effect for consistency scores might contribute to the small differences.

## *What do qualitative analyses reveal when one word is correct and the other in error?*

Only 9% of the words were said correctly in one trial and in error in the other: 70% of those errors were age-appropriate and 19% were delayed developmental errors. Table 7 shows that only 11% of these errors were atypical (e.g. backing). Perhaps consistently used phonological patterns that are about to be or were recently, suppressed are more likely to alternate between developmental errors and correct production. Many words seemed vulnerable to inconsistency, being multi-syllabic with complex syllable structure. Waring (2019) argued that the executive function of inhibition plays an important role in suppression of error patterns. Children may fail to inhibit a familiar phonological pattern when marking an emerging contrast (e.g. alveolar vs velar) in words with a phonologically complex structure (e.g. *kangaroo* [tændəɹʉː]).

A similar explanation might account for the 17 examples of assimilation, usually suppressed by 3;0 years (e.g. *octopus* [ɔpəpʊs]). Dodd (1995, p. 43) noted that 'to cope with new, complex and polysyllabic words' children in a longitudinal study reverted to assimilation errors (e.g. *hippopotamus* [mɪpʌmɔpʌmʊs]). While larger longitudinal studies are needed (Macrae & Sosa, 2015; Martikainen et al., 2019), current evidence suggests that when one word is correct and the other in error, the trial should not be counted for identification of inconsistent phonological disorder.

### What are the differences in phoneme production when a word is pronounced differently across trials?

Inconsistency might, then, be better identified only when a word's two productions contained different errors. There were only 143 examples from all children. Words most vulnerable to inconsistent production had complex word structures. Word frequency and neighbourhood density had less effect, perhaps reflecting sparse Australian word frequency data and that multisyllabic words have few/no neighbours (Rajaram, 2021). While other studies have reported that 'variability' is associated with frequency and neighbourhood density (Jones, 2020; Sosa & Stoel-Gammon, 2012), their data sets have focused on younger participants (12–48 and 24–29 months) and predominantly monosyllabic words. Both word structure (Macrae, 2013) and word length (Sosa, 2015) have previously been associated with increased variability.

Word pairs with different errors were categorised into three types of inconsistency. Most of the word pairs (56%) were in the first category. This category of inconsistency occurred when a segment was in error but with different developmental errors in a word pair: *witch* [wɪt] [wɪts] stopping/fronting of /tʃ/. *Screwdriver* as [skɹʉ:dɑevə] [skʉ:dɹɑevə] demonstrated a child's ability to pronounce the target clusters within the phonetic context of a word but difficulty correctly planning both in one production. These examples suggest that online planning of phoneme sequences allows selection of alternative developmental errors for segments. There may also be a limit on the number of segments a child can include in a phonological plan.

A second category of inconsistency identified word pair differences due to known non-developmental error patterns associated with SSD. Substitution of clusters with an affricate or fricative contributed more than half of these errors: *strawberries* [ʃo:bewi:z] [sto:bewi:z], *screwdriver* [ʃʉ:dɹɑevə] [dʒʉ:dʒɑevə]. Although this error pattern was recently identified in 10% of TD three-year-olds (Holm et al., 2021), normative studies rarely report it. Affrication is, then, classed with atypical errors, like backing, as a marker of atypical phonological development (Khan & Lewis, 2015). These two categories of inconsistency indicate that the phonological assembly stage of Levelt et al.'s (1999) model incorporates developmental and atypical phonological contrasts and constraints that provide the blueprint for building articulatory plans for word production.

The third error category included the remaining 22% of word pairs with different errors. All involved at least one of the pair having word structure oddities: syllable addition; non-developmental deletion; partial replication; phoneme addition; metathesis; cluster creation; unrelated substitutions. There were only 15 word pairs in this category in the data for the youngest children diminishing to 4 in the oldest group, implying a developmental trajectory (see Table A1). The errors affected the sequence of phonemes and syllable structure of words as opposed to articulatory distortions or phonological patterns. These errors, then, plausibly reflect poor phonological planning of the sequence of phonemes in words.

### Do qualitative analyses provide evidence for explanations of TD children's inconsistency?

Both Jones (2020) and Macrae and Sosa (2015) suggest that incomplete or inaccurate phonological representations contribute to inconsistent word production. The inconsistency documented in this paper, however, provides counter evidence. For a word to be produced correctly in one of the trials suggests there is an intact underlying phonological representation for the 9% of word pairs that were correct in one trial but in error in the other. Nor can errors occurring only on one trial be easily classed as articulatory since most errors reflected developmental phonological error patterns that are age appropriate. As Levelt et al. (1999) proposed, speech production seems to include a stage where the sequence of phonemes is selected before articulation is programmed.

The data raises the question of whether an incomplete or inaccurate phonological representation can give rise to two different productions of a word. The 9% of word pairs with two different surface speech errors fell into three categories: more than half were alternative developmental errors; 22% had an atypical error pattern in one trial and often a developmental error in the other. Therefore, most errors could be accounted for by phonological patterns. Few word pairs could be accounted for by an incomplete representation. Less than a quarter of word pairs with two different surface speech errors were characterised by odd word and syllable structure errors. We suggest that this category reflects selection and sequencing errors in a phonological plan.

### *Limitations*

The consistency assessment included only 15 predominantly multisyllabic words. The assessment is in development and has yet to be shown to identify cases of CAS or IPD, although clinical data are currently being collected. The age range of the normative sample was limited to preschool children, and data are needed for a wider age range. Typically developing children were reported to have cognitive, linguistic and speech motor skills within normal limits, but these were not specifically assessed. Nevertheless, given the variable reports of inconsistency for this age group, the study provides important normative data.

## Conclusion

The data showed that children aged 36–60 months most often produced words accurately and consistently (57%) or with both productions of a word having the same error (25%). The few words said differently implies that preschool children have mastered the ability to name a picture, planning a consistent sequence of phonemes from their mental representation to denote a concept shared by adults. This ability is over 80% reliable and words more prone to inconsistency are multisyllabic with complex syllable structure. Further research comparing the types of inconsistent surface speech errors made by TD children, those diagnosed with inconsistent phonological disorder and those with childhood apraxia of speech would evaluate models of speech production in child language acquisition.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Australian National Health and Medical Research Council (Canberra, AU) https://app.dimensions.ai/details/grant/grant.7878989 [GRANT_NUMBER: GNT1153614].

## ORCID

Alison Holm http://orcid.org/0000-0002-9323-8055
Sharon Crosbie http://orcid.org/0000-0002-3873-7468
Angela Morgan http://orcid.org/0000-0003-1147-7405
Barbara Dodd http://orcid.org/0000-0002-2008-6409

## References

American Speech-Language-Hearing Association. (n.d.) Childhood apraxia of speech (*Practice Portal*). Retrieved October 14, 2021, from https://www.asha.org/practice-portal/clinical-topics/acquired-apraxia-of-speech/

Australian Bureau of Statistics (ABS) (2016) Technical Paper: Socio-Economic Indexes for Areas (SEIFA) 2016 (2016·0·55·001). Canberra: ABS. https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001

Bååth, R. (2010). *ChildFreq: An online tool to explore word frequencies in child language*. LUCS Minor, 16, Lund University. Retrieved October 14, 2021, from https://portal.research.lu.se/ws/files/6190321/1776717.pdf

Ball, M. (2016). *Principles of clinical phonology: Theoretical approaches*. Taylor & Francis. https://doi.org/10.4324/9781315670980

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., & Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *PloS One*, *15*(4), e0231484. https://doi.org/10.1371/journal.pone.0231484

Bartolotti, M., Chhabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighbourhood densities. *PLoS One*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267–304. https://doi.org/10.1177/0142723704045681

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders*, *39*(3), 303–324. https://doi.org/10.1080/13682820310001625589

Bürki, A. (2018). Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review*, *25*(December, 2018), 1973–2004. https://doi.org/10.3758/s13423-017-1423-4

Burt, L., Holm, A., & Dodd, B. (1999). Phonological awareness skills of 4-year-old British children: An assessment and developmental data. *International Journal of Language and Communication Disorders*, *34*(3), 311–335. https://doi.org/10.1080/136828299247432

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition*, *42*(1–3), 261–286. https://doi.org/10.1016/0010-0277(92)90045-J

Crosbie, S., Holm, A., & Dodd, B. (2021). Core vocabulary intervention. In A. Williams, S. McLeod, and R. McCauley (Eds.), *Intervention for speech sound disorders in children* (2nd ed., pp. 225–249). Brookes.

Delaney, A., & Kent, R. (2004, November 18-20). *Developmental profiles of children diagnosed with apraxia of speech* [Poster presentation]. *American Speech-Language-Hearing Association Convention*. Philadelphia, PA, United States.

Docherty, G., & Mendoza-Denton, N. (2012). Speaker-related variation - Sociophonetic factors. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Oxford handbook of laboratory phonology* (pp. 43–60). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199575039.013.0004

Dodd, B. (1995). Children's acquisition of phonology. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorders* (pp. 21–48). Whurr.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology*. Pearson.

Dodd, B., & McCormack, P. (1995). A model of speech processing for differential diagnosis of phonological disorders. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorders* (pp. 65–89). Whurr.

Duffy, J. R., Utianski, R. L., & Josephs, K. A. (2021). Primary progressive apraxia of speech: From recognition to diagnosis and care. *Aphasiology*, *35*(4), 560–591. https://doi.org/10.1080/02687038.2020.1787732

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, *30*(Pt 2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*(2), 419–439. https://doi.org/10.2307/412864

Grunwell, P. (1982). *Clinical phonology* (2nd ed.). Croom Helm.

Haley, K. L., Jacks, A., & Cunningham, K. T. (2013). Error variability and the differentiation between apraxia of speech and aphasia with phonemic paraphasia. *Journal of Speech, Language and Hearing Research*, *56*(3), 891–905. https://doi.org/10.1044/1092-4388(2012/12-0161 )

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(2), 155–184. https://doi.org/10.1080/07268609708599550

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language and Communication Disorders*, *42*(4), 467–486. https://doi.org/10.1080/13682820600988967

Holm, A., Sanchez, K., Crosbie, S., Morgan, A., & Dodd, B. (2021). Is children's speech development changing? Preliminary evidence from Australian English-speaking three-year-olds. *International Journal of Speech-Language Pathology*, 1–10. https://doi.org/10.1080/17549507.2021.1991474

Ingram, D. (1976). *Phonological disability in children*. Edward Arnold.

Jones, S. D. (2020). Accuracy and variability in early spontaneous word production: The effects of age, frequency and neighbourhood density. *First Language*, *40*(2), 128–150. https://doi.org/10.1177/0142723719894768

Kehoe, M., & Cretton, E. (2021). Intraword variability in French-speaking monolingual and bilingual children. *Journal of Speech, Language, and Hearing Research*, *64*(7), 2453–2471. https://doi.org/10.1044/2021_JSLHR-20-00558

Khan, L. M. L., & Lewis, N. (2015). *Khan-Lewis phonological analysis* (3rd ed.). PsychCorp.

Laganaro, L. (2019). Phonetic encoding in utterance production: A review of open issues from 1989 to 2018. *Language, Cognition and Neuroscience*, *34*(9), 1193–1201. https://doi.org/10.1080/23273798.2019.1599128

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Maassen, B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. *Seminars in Speech and Language*, *23*(4), 257–266. https://doi.org/10.1055/s-2002-35804

Macrae, T. (2013). Lexical and child-related factors in word variability and accuracy in infants. *Clinical Linguistics & Phonetics*, *27*(6–7), 497–507. https://doi.org/10.3109/02699206.2012.752867

Macrae, T., & Sosa, A. V. (2015). Predictors of token-to-token inconsistency in preschool children with typical speech-language development. *Clinical Linguistics & Phonetics*, *29*(12), 922–937. https://doi.org/10.3109/02699206.2015.1063085

Martikainen, A. L., Savinainen-Makkonen, T., & Kunnari, S. (2019). Intra-word consistency and accuracy in Finnish children aged 3–6 years. *Clinical Linguistics & Phonetics*, *33*(9), 815–830. https://doi.org/10.1080/02699206.2019.1576770

McCabe, P., Thomas, D., Murray, E., Crocco, L., & Madill, C. (2017). *Rapid syllable transition treatment – ReST*. The University of Sydney. Retrieved October 14, 2021 from https://rest.sydney.edu.au/

McCormack, P., & Dodd, B. (1998, September 4-6). *Is inconsistency in word production an artefact of severity in developmental speech disorders?* [Poster presentation]. *Child language seminar*.s Sheffield, UK.

Ozanne, A. (2005). Childhood apraxia of speech. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorders* (pp. 71–82). Whurr.

Rajaram, M. (2021). Phonological neighborhood measures and multisyllabic word acquisition in children. *Journal of Child Language* *49*(1), 197–212. https://doi.org/10.1017/S0305000920000811

Seunghee, H. (2020). Variability and inconsistency in children with and without speech sound disorders. *Communication Science and Disorders*, *25*(2), 431–440. https://doi.org/10.12963/csd.20714

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in Childhood Apraxia of Speech. *Clinical Linguistics and Phonetics*, *26*(5), 445–482. https://doi.org/10.3109/02699206.2012.655841

Smith, N. (1973). *The acquisition of phonology*. Cambridge University Press.

Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology*, *24*(1), 24–35. https://doi.org/10.1044/2014_AJSLP-13-0148

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech Language and Hearing Research*, *55*(2), 596–608. https://doi.org/10.1044/1092-4388(2011/10-0113 )

Verspoor, M., Lowie, W., & de Bot, K. (2021). Variability as normal as apple pie. *Linguistics Vanguard*, *7*(s2), 20200034. https://doi.org/10.1515/lingvan-2020-0034

Vogel Sosa, A., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*, *33*(1), 31–50. https://doi.org/10.1017/S0305000905007166

Waring, R. (2019). *An investigation of the cognitive-linguistic profile of children with phonological delay and phonological disorder* [Unpublished doctoral dissertation]. The University of Melbourne. http://hdl.handle.net/11343/233288

Zarifian, T., Ahmadi, A., & Ebadi, A. (2020). Development and measurement of psychometric properties of the Persian test of speech consistency in children with typical development. *Applied Neuropsychology: Child*, 1–9. Advance publication. https://doi.org/10.1080/21622965.2020.1786831

# Appendix

**Table A1.** Data for four age groups (36–60 months): All occurrences of both words in error.

| | | | | | |
|---|---|---|---|---|---|
| **36–42 months**: 56 words with different error forms from *n* = 24 children | | | | | |
| *Differences between word pairs due to developmental errors*: 30/56 = 54% | | | | | |
| fwɔ | fwɔg | bækjʉ:mki:nə | bækjʉ:mkli:nə | askɹʉ:dɑevə | skʉ:dɹɑevə |
| ɹɔg | fɹɔk | vækɹʉ:mkli:nə | vægɹʉ:mkli:nə | skwʉ:dɹɑevə | skʉ:dɹɑevə |
| bɔg | fɔg | bækjʉ:kli:nə | bækʉ:kli:nə | sɹʉ:dɹɑevə | sɹʉ:ɹɑevə |
| flɔg | fwɔg | l nbɔks l | nbɔts | skwʉ:dwɑevə | skwʉ:dɹɑevə |
| sɹo:beɹi:z | sɹo:beɹi: | l mpɔks l | mpɔk | fwɪŋkəl | fwɪŋkəlz |
| sto:bewi:z | do:bewi:z | l nbɔks l | ntbɔks | bɪkəlz | bɪŋkəlz |
| sto:bi:z | stɹo:bi:z | l nbɔks l | bɔk | ɹɪŋkəl | sɹɪŋkəl |
| so:bewi:z | do:bewi:z | ʃi: | tʃi: | kæɹʉ: | kæwʉ: |
| tɹo:bɹi: | do:bwi:z | efənt | eləfən | slæʃ | tlæʃ |
| sto:bewi:z | stɹo:bewi:z | kwæɔd | kwæɔ: | | |
| *Differences in word pairs involving an atypical error*: 11/56 = 20% | | | | | |
| astɹo:bwi:z | ʃo:bɹi:z | tʃo:bɹi: | sto:bewi: | flæʃ | flæs |
| tʃo:bɹi: | sto:bewi: | pɹɪŋkəlz | fɹɪŋkəlz | kɹʉ:dɹɑevə | kɹʉ:gɹɑevə |
| stɹo:bi:z | gɹo:bi:z | fwɪŋkəlz | kwɪŋkəlz | fɹʉ:dɹɑevə | fɹʉ:dɹɑe:ə |
| tʃo:bɹi:z | fɹo:bɹi:z | | | | |
| *Differences in word pairs due to syllable and word structure errors*: 15/56 = 27% | | | | | |
| æpsəpʊs | ɔptəpʊs | spɹɪŋkə:s | spwɪŋkʊs | hɔtəkɔtə | hɔli:kɔtə |
| ɔtəpʊspʊs | ɔtəpʊs | spwɪŋkʊzə | pwɪŋkəw | heli:kɔftə | heli:kɔktə |
| ɔpəpʊs | ɔptəpʊs | bæjʉ:ti:mə | bækjʉ:mti:nə | ej tɔkwə | ewi:tɔktə |
| kæŋgɹʉ: | kænjʉ: | bɹækjʉ:ki:nə | bækjʉ:mki:nə | hewəkɔtə | hewəklɔtə |
| eflənt | efənt | dʉ:dɹɑevə | nʉ:dɹɑeʃə | hepətɔktə | hɔkdətelə |
| **43–48 months**: 38 words with different error forms from *n* = 24 children | | | | | |
| *Differences between word pairs due to developmental errors*: 23/38 = 61% | | | | | |
| afɔg | fwɔg | skʉ:dɹɑevə | ɹʉ:dɑevə | bɪŋkəlz | pɪŋkəlz |
| aefənt | ewəfənt | stʉ:dɹɑevə | stɹʉ:dɹɑevə | pɹɪŋkəlz | spɹɪŋgəlz |
| pæs | plæs | stʉ:dɑevə | dʉ:dɑevə | pɪŋkəlz | pɹɪŋkəlz |
| bæs | bæʃ | tɹɔbwi: | tɹɔbɹi:z | fɔdi: | fɹɔdi: |
| splæs | spæs | stɹɔbewi:z | stwɔbewi:z | klæɔs | kæɔs |
| j ntʃbɔks j | ntbɔks | wɔbewi:z | stɔbewi:z | heli:tɔtə | hedɔtə |
| l nsbɔks l | bɔks | bækjʉ:mkwi:nə | bækjʉ:mtwi:nə | heli:kɔktə | heli:tɔktə |
| *Differences in word pairs due to atypical errors*: 8/38 = 21% | | | | | |
| sɔbi:z | tʃɔbɹi:z | bægʉ:gi:nə | vægjʉ:ngi:nə | li:p | si:p |
| pjɔbɹi:z | sjɔbɹi:z | gækjʉ:mki:nə | bækjʉ:ki:nə | wɪts | wɪθ |
| fʉ:fɑevə | fʉ:dɑevə | spɹɪgəlz | spɹɪnəlz | | |
| *Differences in word pairs due to syllable and word structure errors*: 7/38 = 18% | | | | | |
| hewi:tɔktə | hewi:tɔtə | ʃʉ:dɹɑevə | dʒʉ:dʒɑevə | pwɪŋkəlz | pətʃɪŋkəlz |
| hɔktətelə | hetətɔktə | spwæs | ɸpwæs | pɹɪŋkəlz | ɸpwɪŋkəlz |
| æzəkɔptə | hewəkɔptə | | | | |
| **49–54 months**: 21 words with different error forms from *n* = 24 children | | | | | |
| *Differences between word pairs due to developmental errors*: 10/21 = 48% | | | | | |
| w ntʃbɔk w | ntʃbɔks | tæŋgəwʉ: | kæŋgəwʉ: | skʉ:dɹɑevə | skʉ:bɹɑebə |
| spæʃ | spwæʃ | pækwʉ:mti:nə | fætwʉ:mtli:nə | stʉ:dɹɑevə | skʉ:dɹɑevə |
| spwæʃ | spwæs | wækjʉ:mkli:nə | wækjʉ:mkwi:nə | pɹɪŋkəlz | bwɪŋkəlz |
| stɹɔbewi:z | stɔbewi:z | | | | |
| *Differences in word pairs due to atypical errors*: 6/21 = 29% | | | | | |
| kwæɔs | kwæɔdz | dækjʉ:mtli:nə | lætjʉ:m tli:nə | kɹʉ:zdɹɑeə | kɹʉ:zgɹɑezə |
| heji:tɔtə | eji:dɔtə | sɪŋkəlz | sɪ:əlz | sʉ:dɹɑevə | sʉ:gɑevə |
| *Differences in word pairs due to syllable and word structure errors*: 5/21 = 24% | | | | | |
| ɔptəpʊs | heli:pʊts | kwæpʉ:kwi:nə | kwæŋkʉ:nkwi:mə | æɔnsənt | æɔsənt |
| kɔjəkɔwə | kɔkə | tɹɔbi:bewi:z | bwɔwi:z | | |

(*Continued*)

**Table A1.** (Continued).

| | | |
|---|---|---|
| **55–60 months**: 28 words with different error forms from *n* = 24 children | | |
| *Differences between word pairs due to developmental errors*: 17/28 = 61% | | |
| afɔg fwɔg | skʉ:tɑeə skɹʉ:tɑeə | skʉ:dɹɑevə stʉ:dɑevə |
| spwɪŋkəl spɹɪŋkəl | wɪt wɪts | helkɔtə heli:kɔtə |
| aspɪŋkəlz spwɪŋkəlz | splæs plæʃ | tæŋgəwʉ: kæŋgəjʉ: |
| stɹɔbewi:z sɔbebi:z | klæɔ kwæɔ | bækjʉ:kli:nə bækjʉ:ki:nə |
| ʃɹɔbewi:z ʃɔbewi:z | ewəbən ewəpən | heləkɔpə heləkɔ:ə |
| *Differences between word pairs due to atypical errors*: 7/28 = 25% | | |
| ʃɔbewi:z stɔbewi:z | skʉ:sɑevə skʉ:zɑevə | fɪŋkəlz swɪŋkəlz |
| stabeɹi: stɔbeɹi: | skʉ:dɹɑevə ʃʉ:dɹɑevə | fɐg fɔg |
| ɔwi:kɔtə hewəkɔtə | | |
| *Differences in word pairs due to word and syllable structure errors*: 4/28 = 14% | | |
| ɔptəkʊpʊs ɔptəkʊs | skɹɪŋgəl skɹɪskɹəl | bækjʉ:piə **v**ækjʉ:vi: |
| ɔpʊ: ɔpəpʊt | | |

[a]2 examples