ASHA American Speech-Language-Hearing Association

LSHSS

Research Article

Is Knowing Words Enough? Assessing Vocabulary in Children Who Are Deaf or Hard of Hearing

Leo Evans,[a] Emily A. Lund,[b] and Krystal L. Werfel[a]

[a] Boys Town National Research Hospital, Omaha, NE [b] Texas Christian University, Fort Worth

ARTICLE INFO

Article History:
Received May 2, 2024
Revision received August 21, 2024
Accepted September 29, 2024

Editor-in-Chief: Kelly Farquharson
Editor: Lisa S. Davidson

https://doi.org/10.1044/2024_LSHSS-24-00046

ABSTRACT

Purpose: Vocabulary skills in children are typically measured with norm-referenced assessments of receptive and expressive vocabulary. Language sample analysis is an alternative method of examining vocabulary actually produced in communicative events and may be better suited to exposing subtle vocabulary weaknesses. Here, we examine the relationship between norm-referenced vocabulary testing and language sample analysis in preschool children, both children with typical hearing (CTH) and children who are deaf or hard of hearing (DHH).
Method: Expressive and receptive vocabulary and spontaneous language samples were collected. Language samples were analyzed for complexity (mean length of utterance in words [MLUw]) and variability (number of different noun and verb types produced).
Results: CTH had significantly higher scores on both expressive and receptive norm-referenced tests and produced sentences with greater syntactic complexity and semantic variability. Relationships between expressive test scores and MLUw were seen in both groups; the number of noun/verb types produced was related for children who are DHH only. Receptive vocabulary was not related to spontaneous spoken language for CTH. Receptive vocabulary was significantly related to MLUw and noun/verb types for children who are DHH. However, when the DHH group was subdivided into performance above and performance below the 50th percentile, relationships held only for the group with below-average performance.
Conclusions: Results suggest that single-word vocabulary, norm-referenced measures indicating performance above the 50th percentile may not be sufficient to capture nuanced difficulties with vocabulary in children who are DHH. For children who are DHH, performance in the "range of normal" on a norm-referenced test may not capture ways in which their language performance differs from that of peers with typical hearing.

A large vocabulary affords a child precise and accurate communication. Having a broad range of words at their disposal enables them to convey wants, needs, ideas, and feelings in an effective manner, diminishing the chance of misunderstandings and promoting more nuanced conversations. Greater vocabulary knowledge in childhood has been found to promote better language outcomes, including increased syntactic competence, greater reading comprehension, and more robust narrative skills (Colozzo et al., 2011; Law & Edwards, 2015; Lee, 2011; Rvachew & Grawburg, 2006), as well as overall higher academic and vocational outcomes (Armstrong et al., 2017).

But what does it mean to know a word? Minimally, knowing words entails knowledge of their lexical configuration, that is, the facts we know about a word (e.g., spelling, pronunciation, meaning; Leach & Samuel, 2007). Aspects of lexical configuration, phonological forms and meanings, are straightforward to test and form the basis of many vocabulary assessment methods (Brownell, 2011; D. M. Dunn, 2019) designed for approximating the breadth of a child's vocabulary or the sheer number of words known (Webb, 2013). In typical assessments used in the

Correspondence to Leo Evans: lee.evans@boystown.org. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

United States, such as the Peabody Picture Vocabulary Test–Fifth Edition (PPVT-5; D. M. Dunn, 2019) and the Expressive One-Word Picture Vocabulary Test–Fourth Edition (EOWPVT-4; Brownell, 2011), test administrators either say the word (i.e., provide its phonological form) and ask the test taker to find the picture matching the word or provide a picture and ask the test taker to say the word corresponding to that picture. Vocabulary breadth is an index of how well children can link the phonological form of a word and its meaning/referent. Linking of form and meaning is a necessary step in vocabulary development, as it sets the stage for word recognition and use (Webb, 2013). Using a word, however, may require a deeper type of knowledge and indicates the depth of a child's vocabulary (Anderson & Freebody, 1981; Nation, 2001; Webb, 2013). Using a word entails accessing it and correctly implementing it in a communicative exchange. Producing words in conversations requires a deeper, more nuanced level of word knowledge than simple word recognition as it requires understanding the meaning of a word in context to appropriately fit it into the larger message of the sentence or conversation as well as the grammatical structure of its sentence. Under such constraints, children may not be able to utilize words they might recognize in less challenging contexts. Producing words in conversations is one way a child can exhibit greater depth of vocabulary knowledge. Assessing utilized vocabulary would require a method in which a child engages in discourse (narrative or conversational) in order to identify which words they actually produce in a given communicative context.

As children learn a spoken language largely by listening, imitating, and attaching meaning to the word forms they encounter, hearing is vital to this process. Children who are deaf or hard of hearing (DHH) are particularly at risk in terms of oral language development, including vocabulary, given the potential for limited auditory access to spoken language (Walker et al., 2015), and often enter school with less vocabulary knowledge than children with typical hearing (CTH; Lund & Douglas, 2016; Tomblin et al., 2015; Werfel et al., 2022). Some DHH children learning a spoken language who are identified early and receive support in the form of hearing devices can have outcomes comparable to those of CTH in preschool (Geers et al., 2016; Walker et al., 2022). However, this outcome is often determined utilizing norm-referenced language tests, which evaluate language skills (i.e., vocabulary or grammar) by examining responses to a small number of specific linguistic tasks or vocabulary questions selected for their utility in quickly identifying where a child's ability lies compared to same-age peers. Such tests are designed to quickly identify vocabulary or grammatical deficits. For vocabulary assessment in particular, the trade-off is that norm-referenced vocabulary assessments only investigate surface-level knowledge of phonological forms and referent meanings. This method of assessment differs substantially from the daily use of language for communication, which requires the child to have deep enough word knowledge to produce the word under the challenges of a conversation. That is, they must produce the correct word, along with many other words, in correct grammatical order to effectively communicate their thoughts. Given their emphasis on surface-level knowledge, it is not certain that a child's performance on norm-referenced vocabulary assessment tasks is an accurate portrayal of their vocabulary skills in speaking (Ebert & Scott, 2014; Sénéchal et al., 2008; Souza & Cáceres-Assenço, 2021).

Despite the reliance on norm-referenced assessments of language long being viewed as unsatisfactory (M. Dunn et al., 1996; Spaulding, 2012; Spaulding et al., 2006) due to its lack of ecological validity, it remains a common assessment mode within educational and clinical settings. A second less common method is the collection and analysis of natural language samples (e.g., the spontaneous spoken language produced by a child during a conversation or telling of a story, often elicited via specific prompts). Language sample analysis, although more time consuming, may be better suited to exposing language weaknesses (Werfel & Douglas, 2017) because, unlike norm-referenced tests that do not probe the actual use of language during ongoing communication, analyses of spontaneous spoken language samples give a more acute view of what language skills a child is able to use when engaging in their everyday environment. Examining the spontaneous spoken language samples of children can reveal weaknesses that are not evident on norm-referenced tests. As eligibility for services is often dependent on the evaluation of norm-referenced tests (Nowland, 2009; Skahan et al., 2007), children with hearing loss who perform well on norm-referenced tests but struggle in real life may not receive services they need to succeed in the classroom (Werfel & Douglas, 2017).

Differences in Vocabulary Assessment Style

Work involving the comparison of language sample analysis and norm-referenced language assessments focusing on syntactic performance (i.e., children's performance on grammatical assessments of morphosyntax compared to their productions of syntactic structures during spontaneous speech) has found language sample analysis well suited to detecting atypical language development (M. Dunn et al., 1996; Ramos et al., 2022; Werfel & Douglas, 2017). A similar evaluation of lexical-semantic knowledge (e.g., vocabulary) across assessment styles is warranted. Measuring vocabulary is always an approximation, as it

would be impossible past a certain point to precisely delineate all the words a child knows. Two methods of assessing vocabulary are norm-referenced assessments and analyses of spontaneous spoken language. These methods differ not only in the purpose and style of assessment but also in the cognitive and linguistic demands placed on the child.

Norm-referenced assessments of vocabulary occur in a highly controlled manner. They are administered in a consistent way to reduce administrator bias and make results easily comparable. These tests have been constructed to include a limited number of highly predictive items that have been administered to a large number of individuals to provide typical age-based response ranges. By comparing performance on these items to age-based normative scores, a clinician can determine if a child is on par with, behind, or advanced compared to same-age peers.

For receptive vocabulary tests (e.g., Comprehensive Receptive and Expressive Vocabulary Test [Wallace & Hammill, 2002], PPVT-5 [D. M. Dunn, 2019], Receptive One-Word Picture Vocabulary Test [Martin & Brownell, 2011]), children typically see a set of pictures and are asked to point to the one that corresponds to the word they hear (e.g., "find giraffe" in a set of pictures including a giraffe, a zebra, a rhinoceros, and a cheetah). This method, however, allows a child to guess a referent correctly despite not fully knowing the word, and guessing as a strategy is often encouraged during testing. Providing the correct response on such an assessment does not, therefore, guarantee that the child could produce the word when speaking or even recognize the word in a conversational encounter.

Single-word expressive vocabulary is typically measured by showing a series of pictures and asking the child to label the object seen (e.g., Comprehensive Receptive and Expressive Vocabulary Test [Wallace & Hammill, 2002], EOWPVT-4 [Brownell, 2011], Expressive Vocabulary Test–Third Edition [Williams, 2019]). Another common expressive task is the solicitation of synonyms of words. Although this type of task does reveal whether the child can produce the target word, a more challenging task than word recognition, accurate production in this context, is not a guarantee that the child would produce the word unprompted during an actual conversation.

In order to tap into vocabulary depth, an alternative assessment method is needed that increases the demands placed on the child. One such alternative method is to estimate children's vocabulary using natural language samples of spontaneous spoken language from a more naturalistic context. The purpose of a language sample is to discover the words a child produces in actual communicative events (Evans, 1996; Klee, 1985). Unlike a norm-referenced test, there is no limit to the words tested as the child is free to say whatever words come to mind when speaking.

The cognitive and linguistic demands of norm-referenced tests also differ from those of spontaneous spoken language. For many norm-referenced assessments of vocabulary, the cognitive load placed on a child is reduced as only one word is given at a time, with no syntactic demands, and there are almost always visual supports in the form of pictures (e.g., Peabody Picture Vocabulary Test–Fourth Edition [PPVT-4; L. M. Dunn & Dunn, 2007], PPVT-5 [D. M. Dunn, 2019]). The testing situation is self-paced and predictable, as all questions are in the same format. In contrast, common methods to elicit spontaneous spoken language samples include free conversation (e.g., unplanned exchanges between a child and an experimenter) or more structured contexts such as narratives or expository discourse. Children are learning conversational skills in early childhood (Forrester & Cherington, 2009; Nikolaus & Fourtassi, 2023), and the context of collecting natural language samples from free conversation increases the cognitive load placed on the child in several ways. The questions and answers of the examiner contain multiple words and are syntactically rich. There is little predictability as the child cannot know what the examiner will say next. Sentences come quickly, and there is minimal visual support. Spontaneous spoken language occurs in real communicative events and reveals more of a child's true communicative ability than norm-referenced tests (Evans, 1996; Klee, 1985).

Relationship Between Performance on Norm-Referenced Assessments and Language Samples

Few studies have directly compared the results of norm-referenced tests of vocabulary with vocabulary produced during communicative events. Collectively, findings suggest only a modest relationship between vocabulary test results and spontaneously produced vocabulary in typically developing monolingual preschoolers, with more evidence of a relationship between the two in children with developmental disorders (Condouris et al., 2003; Ebert & Scott, 2014; Sénéchal et al., 2008; Souza & Cáceres-Assenço, 2021; Ukrainetz & Blomquist, 2002). Ukrainetz and Blomquist (2002) reported significant weak-to-moderate relationships between both receptive and expressive norm-referenced vocabulary test scores and the number of different words used in conversations with an experimenter in a sample of typically developing 4- to 6-year-old children. However, despite this overall correlation between norm-referenced assessment performance and

spontaneous spoken language, test performance predicted only 25% of the bottom performers and 25% of the top performers in terms of word use.

No relationship between norm-referenced expressive vocabulary test scores and the number of words produced during a narrative task was observed by Souza and Cáceres-Assenço (2021) in a sample of typically developing 4- to 6-year-old children. Similarly, Senechal et al. (2008) found no relationship between performance on a norm-referenced assessment of expressive vocabulary and narrative performance in a group of typically developing 4-year-old children. These findings from typically developing monolingual preschool children suggest limited alignment between norm-referenced test scores and vocabulary produced in spontaneous spoken language. Ebert and Scott (2014) also did not see correspondence between PPVT scores and the number of different words used in a sample of children aged 6–12 years who had been referred for speech-language evaluations.

A stronger relationship between norm-referenced test performance and spontaneous spoken language has been found in atypically developing children. Condouris et al. (2003) examined the relationship between scores on norm-referenced vocabulary assessments (expressive and receptive) and vocabulary produced during spontaneous spoken language in a group of autistic children between the ages of 4 and 14 years. Participants showed a significant positive relationship between their receptive vocabulary scores (assessed via the Peabody Picture Vocabulary Test–Third Edition [PPVT-3]) and the number of different words they produced during the spontaneous spoken language task. However, the majority of participants performed more than 1 *SD* below the mean on the PPVT-3, so this relationship is evident in children performing more modestly on the norm-referenced measure.

Overall, there is limited evidence that norm-referenced vocabulary assessments should be relied upon as reflections of the vocabulary children produce in real communicative events. In other words, children who perform well on highly controlled and less cognitively demanding norm-referenced tests may not also be those who show the most variability in productive vocabulary in a more challenging communicative event (Condouris et al., 2003; Ukrainetz & Blomquist, 2002). Conversely, some children without any language impairment may perform poorly on the limited items found on the norm-referenced tests due to cultural, economic, or dialectic differences; their typical language skills are evident in conversational contexts (de Villiers & Johnson, 2007; Stockman, 1999, 2000; Washington & Craig, 1992). This does not mean that single-word norm-referenced tests of vocabulary are inaccurate measures of word knowledge but that the purpose for which they were designed was to determine whether a substantial deficit in word knowledge is present when compared to same-age peers. In the identification of such language deficits, they are useful tools, but if they are overextended to serve as proxies for deep word knowledge, evidence for their efficacy is more limited.

For children who are DHH, the additional challenges of navigating a spoken conversation may either reduce their ability to word-find, such that they rely upon fewer, more established words, or only draw on words of which they have the deepest knowledge. Although greater vocabulary breadth is a positive resource to have, some children may nonetheless produce a more modest variety of words in actual conversation. Over time, this may weaken social and learning interactions. In order to prevent this, some children may need support in either accessing or deepening vocabulary knowledge.

Current Project

Comparisons of spontaneous spoken language and norm-referenced vocabulary assessments have yielded mixed results in CTH. Differential performance on these two types of assessments reveals that vocabulary knowledge as measured by norm-referenced tests may not always be equivalent to productive vocabulary in everyday use. Thus, the type of assessment is a critical factor to consider in the measurement of lexical-semantic knowledge in children. Performance on norm-referenced language tests focusing on syntactic ability has been previously found to overestimate grammatical skills in children who are DHH compared to grammatical usage in their spontaneous language samples (Werfel & Douglas, 2017); however, to date, there has been no such investigation of lexical-semantic knowledge. To address this gap in our understanding of language development in children who are DHH, we investigated the relation between performance on norm-referenced receptive and expressive vocabulary assessments and vocabulary produced in spontaneous language samples. We examined this relationship in preschool in CTH and children who are DHH who use listening devices (e.g., digital hearing aids and/or multichannel cochlear implants). Our specific research questions relate to (a) whether performance on norm-referenced vocabulary assessments relates to spoken language variability in spontaneous communicative exchanges in DHH children and CTH and (b) whether a similar pattern is seen in both groups.

Method

Spontaneous language samples, standardized according to the Hadley (1998) protocol, and single-word norm-referenced assessments of expressive and receptive vocabulary

were collected as part of a larger longitudinal study of language development in children with hearing loss (Werfel, 2017; Werfel et al., 2022). All procedures were approved by the institutional review board (IRB) of the University of South Carolina, which served as the IRB of record for this multisite study. Testing took place individually in quiet rooms at the child's home, the child's school, a public library, or the research lab. Assessments were administered by certified speech-language pathologists or trained master's students in speech-language pathology. The total time for all assessments included in the longitudinal study was 1.5–2 hr, occurring in either one or two sessions depending on the needs of the child.

Participants

Eighty-seven children were included in this study. Children ranged in age from 3;9 (years;months) to 4;3 and included CTH and children who are DHH. CTH were recruited locally and had to pass a bilateral hearing screening at or between 20 dB to be included in the study. Children who are DHH were recruited throughout the United States and had a diagnosis of permanent hearing loss by a certified audiologist. Hearing loss in DHH children ranged from moderate to profound, and all used amplification in the form of either cochlear implants or digital hearing aids. All children used spoken English as their primary language (at least 70% of the time), and caregivers reported no additional diagnoses known to impact language development (e.g., Down syndrome, autism). Demographic information for participants is provided in Table 1a, with additional information about children who are DHH in Table 1b. The two groups did not differ in age, $F(1, 86) = 0.282$, $p = .597$, at the time of testing. The mothers of the children who are DHH had a higher level of education than the mothers of CTH, $F(1, 86) = 5.31$, $p = .023$, and the children in the two groups differed on nonverbal intelligence as measured by standard scores on the Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008), $F(1, 86) = 8.57$, $p = .004$. Although CTH had significantly higher scores than children who are DHH, both group means were within the typical range.

Norm-Referenced Vocabulary Tests

Expressive vocabulary was assessed using the EOWPVT-4 (Brownell, 2011). Receptive vocabulary was assessed using the PPVT-4 (L. M. Dunn & Dunn, 2007) or the PPVT-5 (D. M. Dunn, 2019), depending on the assessment date. All children were administered the current version of the PPVT during their testing session. Administration of both measures followed the established procedures set forward in the test protocol. For the EOWPVT-4, children saw a series of colorful pictures and were asked to give the label of the object, action, or concept illustrated. For the PPVT, a series of sets containing four colored pictures were presented, and the children were asked to point to the picture that corresponded with the word said by the examiner.

Spoken Language Samples

In order to measure spontaneous language production, conversational language samples were elicited via the Hadley (1998) protocol. Under this protocol, children converse with the examiner in three 4-min blocks: personal narratives about family and family celebrations such as a birthday party or recent holiday, expository description rules for a favorite game or how to take care of a pet, and narratives retelling a favorite movie or book. Each language sample was approximately 12 min in length, with all four blocks collected sequentially in one session. Sessions were audio- and/or video-recorded for subsequent transcription, and the entirety of the language sample gathered from each child was analyzed. Language sample transcription used SALT programming software (Miller & Iglesias, 2012); procedures are described in detail in Werfel

Table 1. Demographic information.

(a) Demographic information for all participants					
Group	***n***	**Age in months**	**Sex assigned at birth**	**Mother's years of education**	**PTONI**
TH	50	47.84 (1.83)	Male: 20 Female: 31	16.21 (2.56)	112.55 (10.50)
DHH	37	48.06 (1.77)	Male: 19 Female: 18	17.53 (2.67)	103.63 (18.06)
(b) Demographic information for children who are deaf or hard of hearing					
Hearing device	***n***	**Age of identification in months**	**Age of first hearing aid in months**	**Age of first cochlear implant in months**	
Hearing aid	17	14.43 (16.78)	16.73 (16.23)	n/a	
Cochlear implant	20	5.71 (9.67)	8.74 (11.12)	22.9 (15.44)	

Note. PTONI = Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008); TH = typical hearing; DHH = deaf or hard of hearing.

and Douglas (2017) and Werfel (2018). An iterative transcription process by a team including certified speech-language pathologists or advanced speech-language pathology graduate students ensured final transcription agreement of 100%. The first person in the chain made a first-pass transcription from the session recording, then the second person fine-tuned the transcription. Finally, a third member of the group checked the transcript. Any discrepancies were sent back to the initial transcriber, and the process continued until there was full agreement on the transcription. Children's language samples were analyzed to provide metrics of language complexity (i.e., how able was the child in putting words together to produce longer, more involved sentences), volubility (i.e., how much did the children speak), and variability (i.e., how many different words did the children use when conversing). Mean length of utterance in words (MLUw), a metric of language complexity, and the total number of utterances (TNU) and total number of words (TNW), two metrics of productivity volume, were calculated from children's sentences that were fully realized using SALT.

In addition to the TNW produced, other metrics of language variability were also calculated: number of noun types (i.e., the number of unique nouns produced) and verb types (i.e., the number of unique verbs produced). For each child, the list of nouns and verbs produced during spontaneous spoken language was extracted using SALT. The total number of unique nouns and verbs was calculated for all utterances the child produced, including abandoned utterances and mazes. Noun lists were inspected to extract nouns not indicative of a class of objects. Proper names of people (e.g., Sarah) and characters (e.g., Elsa, Olaf) were removed, as were titles of family members (e.g., mom, dad), as at the age of 4 years, it is not clear which children know mom is a noun and which think it is a name. Movie, cartoon, or book titles and names of games were also removed (e.g., Paw Patrol, Chutes and Ladders). Transcripts were consulted if words were ambiguous (e.g., Frozen was removed if used as a title but not if used as a descriptor). Last, singular and plural forms were combined so that only root words were counted as a unique noun (e.g., "cat" and "cats" both counted as tokens of the type "cat"). Lists of verbs were inspected to remove modal and auxiliary verbs, and all tenses were collapsed so that only the main root verb was counted as a unique verb (e.g., "play," "played," and "playing" were all counted as tokens of the type "play").

Data Analysis

To examine group-level differences in norm-referenced language measures and metrics obtained from language samples, a series of one-way analyses of variance were employed. In order to examine relationships between variables within groups, correlational analyses were conducted.

Results

The purpose of this study was to compare performance on norm-referenced tests of vocabulary and vocabulary production during spontaneous spoken language for CTH and children who are DHH, as well as examine the relationship between performance on norm-referenced tests of vocabulary and vocabulary production for both groups.

Group-Level Results

The means, standard deviations, and group comparisons of the norm-referenced assessments of receptive (PPVT) and expressive (EOWPVT-4) vocabulary are presented in Table 2. At the group level, performance on the norm-referenced vocabulary measures differed significantly by hearing status. CTH had significantly higher raw scores, norm-referenced scores, and percentile ranks on both expressive and receptive tests compared to children who are DHH. No significant differences in performance were observed between the children with hearing aids and the children with cochlear implants.

Differences between typical hearing and DHH groups were also observed within spontaneous language samples. Table 3 provides the means, standard deviations,

Table 2. Children's raw, norm-referenced, and percentile results for receptive and expressive vocabulary assessments by hearing status.

Variable	TH	DHH	*F*	η_p^2	HA *n* = 17	CI *n* = 20
PPVT-4 raw scores	85.04 (19.28)	64.03 (25.05)	19.77***	.187	67.53 (28.91)	61.05 (21.56)
PPVT-4 standard	113.57 (13.96)	94.05 (15.2)	38.9***	.311	95.94 (17.15)	92.45 (13.56)
PPVT-4 percentiles	72.96 (25.95)	39.27 (28.71)	33.05***	.278	44.59 (31.06)	34.75 (26.51)
EOWPVT-4 raw scores	61.31 (13.47)	47.62 (16.2)	18.67***	.178	49.18 (16.13)	46.30 (16.56)
EOWPVT-4 standard	116.63 (12.817)	100.89 (16.77)	24.9***	.225	102.41 (16.57)	99.60 (17.26)
EOWPVT-4 percentiles	80.82 (20.37)	52.49 (29.18)	28.81***	.251	55.53 (31.36)	49.90 (27.73)

Note. TH = typical hearing; DHH = deaf or hard of hearing; HA = hearing aid; CI = cochlear implant; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (L. M. Dunn & Dunn, 2007); EOWPVT-4 = Expressive One-Word Picture Vocabulary Test–Fourth Edition (Brownell, 2011).
***$p < .001$.

Table 3. Children's spontaneous spoken language results by hearing status.

Variable	TH	DHH	*F*	η_p^2	HA *n* = 17	CI *n* = 20
MLUw	4.68 (1.25)	3.97 (1.33)	6.57*	.072	3.96 (1.29)	3.97 (1.4)
TNW	429.63 (158.65)	331.24 (153.72)	8.46**	.085	322.53 (145.8)	338.29 (163)
Noun types	41.92 (13.52)	32.55 (17.00)	8.47**	.085	30.53 (15.70)	34.19 (18.20)
Verb types	28.40 (10.15)	23.13 (10.53)	5.74*	.054	20.65 (8.96)	25.14 (11.47)
TNU	90.22 (22.12)	80.55 (25.27)	3.61	.041	77.71 (23.05)	82.86 (27.27)

Note. TH = typical hearing; DHH = deaf or hard of hearing; HA = hearing aid; CI = cochlear implant; MLUw = mean length of utterance in words; TNW = total number of words; TNU = total number of utterances.

*$p < .05$. **$p < .01$.

and group comparisons for the measures derived from children's spontaneous spoken language samples. CTH produced sentences with greater syntactic complexity, indexed by significantly higher MLUw, compared to children who are DHH. Additionally, CTH produced sentences with greater semantic variability, indexed by significantly more unique nouns and verbs, compared to peers who were DHH. CTH produced more words overall compared to children who are DHH, but there was no significant difference in TNU in the spontaneous spoken language of CTH and children who are DHH. No statistically significant differences were observed between children with hearing aids and children with cochlear implants.

Relationships Between Norm-Referenced Measures and Language Samples

Next, we examined individual differences in vocabulary produced during spontaneous spoken language and lexical-semantic knowledge assessed using norm-referenced tests of receptive and expressive vocabulary. A series of Pearson product–moment coefficients were derived to test for correlations between norm-referenced vocabulary test scores and our spontaneous spoken language measures. Percentile scores on norm-referenced tests were used in all calculations, the results of which can be seen in Table 4. For both CTH and children who are DHH, performance on norm-referenced tests of expressive and receptive vocabulary was significantly positively related, as children who performed well on one test also tended to score well on the other. Scores on norm-referenced tests of expressive vocabulary were significantly positively related to narrative MLUw and TNW for both groups, indicating that those who provided more word labels in the norm-referenced testing environment also produced longer, more complex sentences and used more words during the conversation with the experimenter regardless of hearing status. The number of noun types and tokens produced was related to expressive vocabulary scores for children who are DHH only. Children who are DHH who were able to provide more object labels in the testing setting also produced more noun and verb types when speaking; however, no such relationship was evident for CTH.

Performance on the receptive vocabulary assessment was not significantly related to any of the spontaneous

Table 4. Correlations between measures of spontaneous spoken language and percentile scores on vocabulary assessments.

Measure	1	2	3	4	5	6	7
1 EOWPVT-4	—	.777**	.712**	. 049	.690**	.598**	. 603**
2 PPVT-4	*.587***	—	.549**	–.037	.584**	.450**	.447**
3 MLUw	*.295**	*.226*	—	.006	.801**	.650**	.683**
4 TNU	*.046*	*–.098*	*.212*	—	.274	.252	.408*
5 TNW	*.318**	*.168*	*.795***	*.601***	—	.753**	.802*
6 Noun types	*.24*	*.134*	*.685***	*.513***	*.855***	—	.864**
7 Verb types	*.258*	*.085*	*.614***	*.515***	*.862***	*.711***	—

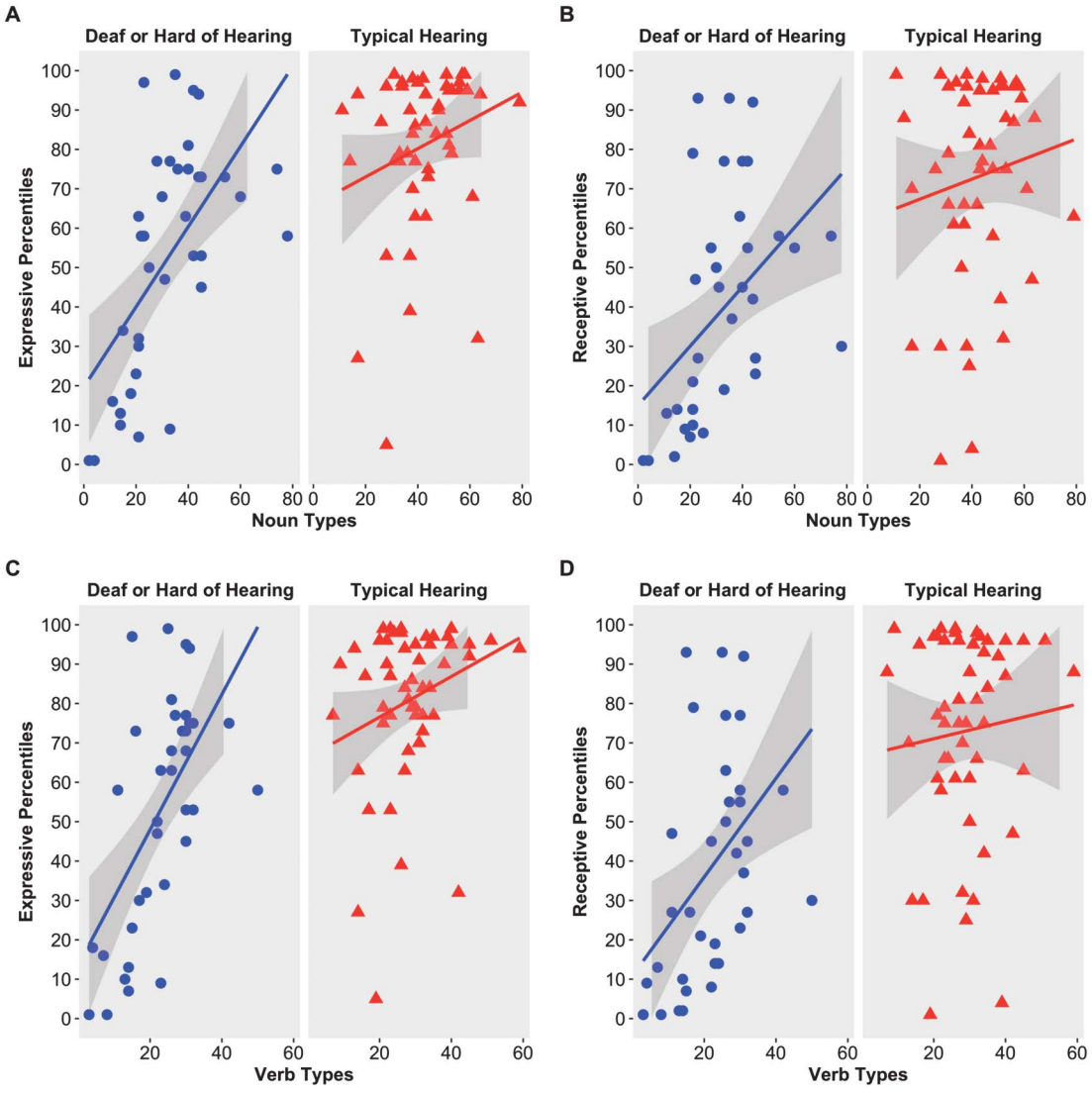
Note. Numbers in italics in the lower left area below the dashes represent the correlations for children with typical hearing, and the numbers above the dashes are correlations for children who are deaf or hard of hearing. EOWPVT-4 = Expressive One-Word Picture Vocabulary Test–Fourth Edition (Brownell, 2011); PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition (L. M. Dunn & Dunn, 2007); MLUw = mean length of utterance in words; TNU = total number of utterances; TNW = total number of words.

*$p < .05$. **$p < .01$.

spoken language metrics for CTH (see Table 4). For CTH, performance on norm-referenced receptive vocabulary tests does not seem predictive of greater sentence complexity or noun verb variety while speaking. Receptive vocabulary scores were significantly related to MLUw and the number of noun and verb types produced for children who are DHH. Children with hearing loss who were able to find referents for object labels in the norm-referenced testing environment were also able to produce longer, more complex sentences containing more noun and verb types in spontaneous conversation. For both the typical hearing and DHH groups, the number of utterances produced was independent of performance on the norm-referenced tests of vocabulary. Correlations are presented in Table 4, and Figure 1 presents the relationship between receptive vocabulary percentile scores and receptive vocabulary percentile scores and noun types produced for both groups.

To further examine the relationship between norm-referenced testing performance and spontaneous spoken language in children who are DHH, the children were

Figure 1. Linear relationships between norm-referenced vocabulary assessments and noun and verb production in spontaneous language samples. (A) Relationship between percentile scores on the Expressive One-Word Picture Vocabulary Test–Fourth Edition (EOWPVT-4) and the number of unique noun types produced during spontaneous spoken language for children who are deaf or hard of hearing (DHH) in blue and children with typical hearing (CTH) in red. (B) Relationship between Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4) percentile scores and the number of unique noun types produced during spontaneous spoken language for both groups. (C) Relationship between percentile scores on the EOWPVT-4 and the number of unique verb types produced during spontaneous spoken language for children who are DHH in blue and CTH in red. (D) Relationship between PPVT-4 percentile scores and the number of unique verb types produced during spontaneous spoken language for both groups.



subdivided into two groups: children who performed at the 50th percentile or above on the EOWPVT-4 and those who performed at the 49th percentile or below. This point was chosen as a halfway point, as children in the first group had scores consistent with the upper half of children their age, and children in the second group had scores consistent with the lower half of children their age. Correlations between norm-referenced test percentiles and spontaneous spoken language measures are shown in Table 5. This subdivision reveals that relationships between norm-referenced tests of vocabulary and vocabulary produced during spontaneous spoken language are driven by children with lower norm-referenced scores, as relationships only hold for children who performed below average on the tests. In other words, for children who performed below the 50th percentile on the norm-referenced tests, test scores align with production variability during spontaneous spoken language. Within the subgroup of below-average performers, those with relatively higher scores also produced more noun and verb types while speaking with the experimenter. For children with hearing loss who had expressive norm-referenced vocabulary test scores at the 50th percentile or above, their scores were not related to spontaneous production; some high scores on norm-referenced tests produced a high number of noun and verb types, whereas others with high norm-referenced scores produced few. The relationship between scores and the production of noun and verb types by percentile group can be seen in Figure 2. Similar relationships can be seen between the expressive vocabulary groups on the PPVT (see Figure 2).

Discussion

Vocabulary, an essential facet of language, is typically assessed in clinical and educational settings via norm-referenced assessments, which are designed to measure the likelihood that a child has a vocabulary deficit. Language sample analysis, although more time consuming, is an alternative, more ecologically valid way to examine children's vocabulary knowledge by evaluating the words they spontaneously produce in conversations. Here, we examined the relationships between results from norm-referenced tests of receptive and expressive vocabulary and vocabulary production during spontaneous spoken language in preschool children who are DHH and preschool CTH. As norm-referenced tests are routinely used in the determination of eligibility for language support services, it is important to determine whether or not they are useful estimates of actual vocabulary usage, a richer metric of vocabulary skills. The findings of this study indicate circumstances in which norm-referenced measures may fail to capture difficulties a child has with language knowledge, namely, that children who score above the 50th percentile on norm-referenced vocabulary measures do not necessarily have high lexical-semantic productivity in conversational language.

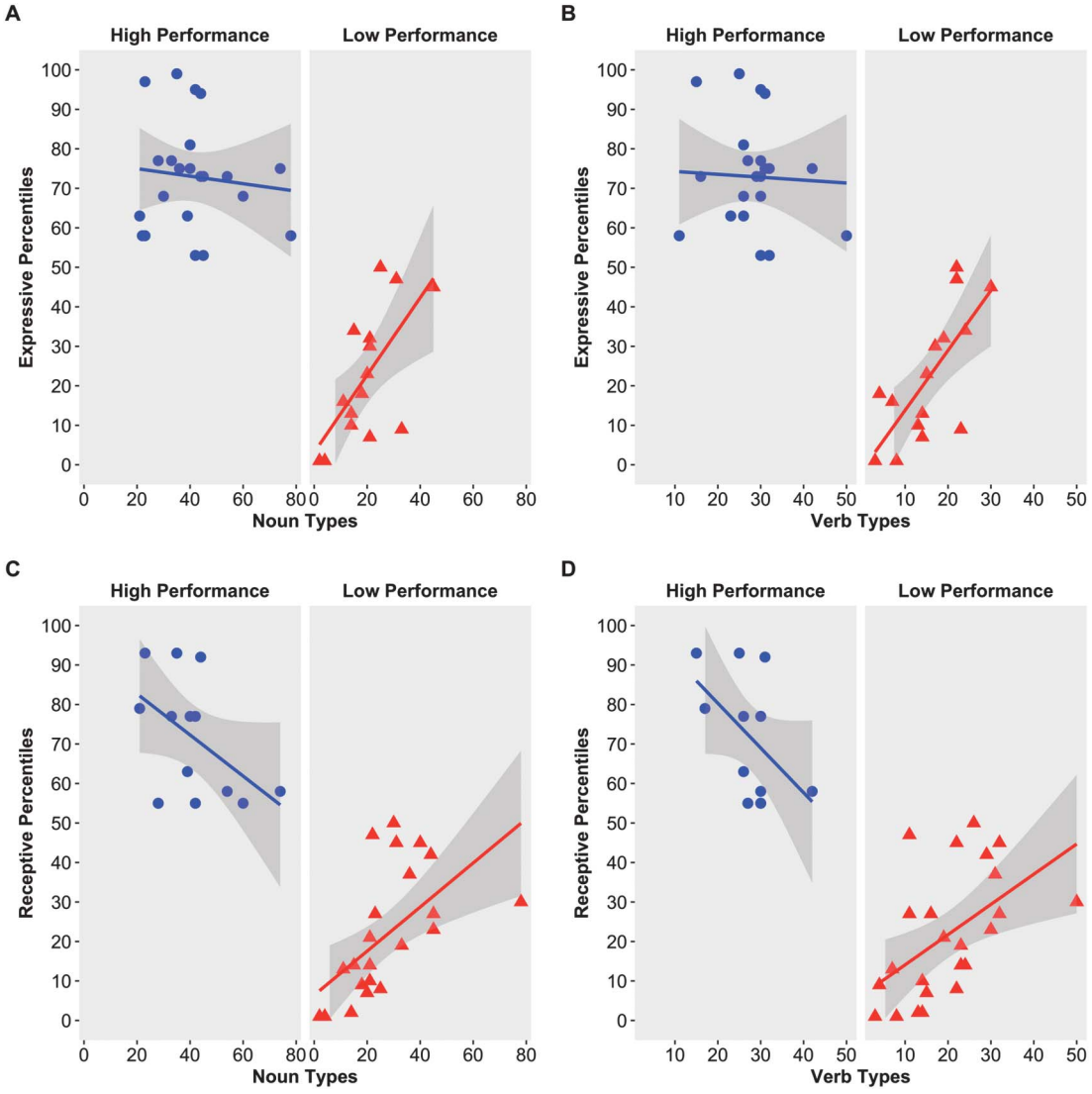
In our sample of preschool children, CTH outperformed children who are DHH on norm-referenced tests of both receptive and expressive vocabulary, indicating greater lexical-semantic knowledge from which to pull during communicative events. Performance on receptive and expressive tests was positively related in both the typical hearing and DHH groups, as most children who did well on one test also did well on the other, indicating that the PPVT and the EOWPVT-4 were tapping into similar underlying vocabulary constructs despite one being rooted in word recognition and the other in labeling. CTH also produced more robust spontaneous spoken language samples when speaking with the examiner; compared to children who are DHH, CTH displayed more syntactic maturity by using more complex sentences, indexed by a longer MLUw, and more varied vocabulary, indexed by a greater variety of unique noun and verb types. CTH and children who are DHH produced similar numbers of intelligible utterances, indicating that group differences in syntax and vocabulary were not likely due to overall reduced productivity on the part of the children who are DHH. The children who are DHH in our sample were engaging with the experimenter to the same degree as peers with typical hearing, but despite this

Table 5. Correlations between measures of spontaneous spoken language and percentile scores on vocabulary assessments in children who are deaf or hard of hearing.

Measure	MLUw	TNW	TNU	Noun types	Verb types
Expressive Vocabulary					
EOWPVT Over 50	.392	.341	–.147	–.104	.076
EOWPVT Under 50	.645**	.730**	.535**	.656**	.723**
Receptive Vocabulary					
PPVT Over 50	*–.179*	*–.236*	*–.246*	*–.104*	*–.047*
PPVT Under 50	*.524***	*.571***	*.055*	*.580***	*.524***

Note. Numbers at the top represent correlations of expressive vocabulary percentiles and the numbers in italics at the bottom represent correlations for receptive percentiles. EOWVT = Expressive One Word Vocabulary Test (Brownell, 2011); PPVT = Peabody Picture Vocabulary Test–4 (L. M. Dunn & Dunn, 2007); MLUw = mean length utterance by words; TNU = total number of utterances; TNW = total number of words.
**$p < .01$.

Figure 2. Linear relationships between performance on norm-referenced vocabulary assessments of expressive (Expressive One-Word Picture Vocabulary Test–Fourth Edition [EOWPVT-4]) and receptive (Peabody Picture Vocabulary Test–Fourth Edition [PPVT-4]) vocabulary and noun production in spontaneous language samples for high and low performers on the EOWPVT-4. (A) Scatter plot of expressive vocabulary percentiles and noun types produced by the expressive percentile group. (B) Scatter plot of expressive vocabulary percentiles and verb types produced by the expressive percentile group. (C) Scatter plot of receptive vocabulary percentiles and noun types produced by the expressive percentile group. (D) Scatter plot of receptive vocabulary percentiles and verb types produced by the expressive percentile group.



engagement, their spontaneous spoken language on average was characterized by simpler, shorter sentences and less varied vocabulary.

Results on Norm-Referenced Assessments Related to Language Samples for Some Children

Examination of individual differences in performance revealed that, for CTH, vocabulary size determined by norm-referenced tests was largely independent of productive vocabulary in spontaneous spoken language; children with higher norm-referenced test scores did not necessarily produce a larger variety of noun and verb words when speaking compared to those with lower scores. Performing well on the tests was not always indicative of utilizing a large number of words to speak more precisely and thus communicating their ideas more effectively, indicating that these two assessment techniques are tapping different levels of lexical-semantic knowledge. This lack of relationship is in line with previous studies of typically developing preschoolers who showed, at best,

modest relationships between the two (Condouris et al., 2003; Ukrainetz & Blomquist, 2002) and not entirely unexpected, as the measures are designed for different assessment purposes. It is worth noticing that, as a group, the CTH in our sample performed very well on both the PPVT and the EOWPVT-4. Almost all in the typical hearing group (92%) scored above the 50th percentile on the expressive measure, and 77% scored above the 75th percentile. For receptive vocabulary, over 80% of the children scored above the 50th percentile, and 53% scored above the 75th percentile. This group of children showed high general word knowledge, although this knowledge did not always translate into greater variability of words produced in conversations. At this age, it appears that norm-referenced test performance reveals little about the variability of words a child will produce when speaking.

For children who are DHH, an overall correspondence was seen between norm-referenced assessment scores and variability in spontaneous spoken language. Children who had higher test scores produced more unique noun and verb tokens in spontaneous spoken language. This aligns with Condouris et al.'s (2003) finding that spontaneous spoken language measures and norm-referenced assessments were positively correlated for autistic children. The children examined by Condouris et al. performed relatively poorly on the norm-referenced vocabulary measures, with most scoring below the 50th percentile. As the strength of association between norm-referenced tests and language samples in our study seemed to be driven by children with relatively poorer performance on the norm-referenced assessment, we examined the children who are DHH in our sample to further determine if relative performance on the norm-referenced measures was also differentially related to spontaneous speech. Of the DHH children in our sample, 62% scored below the 50th percentile on the PPVT, and 38% scored below the 50th percentile on the EOWPVT-4. Therefore, a sizable number of children in our sample performed modestly on these assessments. Within the group of children who displayed less-than-average lexical-semantic knowledge in a norm-referenced test, it was in fact the case that stronger performance on norm-referenced measures coincided with greater variability in spoken language. For children scoring below the mean on norm-referenced tests, test performance was a predictor of spontaneous spoken language performance. For the other children in the DHH group, that is, those who had average or above-average performance on the norm-referenced vocabulary tests, a different pattern emerged. For children scoring above the mean on norm-referenced tests, a higher score did not always coincide with greater variability in spoken language. In fact, some children who scored above the 90th percentile produced relatively few unique noun and verb tokens when conversing. For the children who are DHH scoring above the 50th percentile, test performance was not a predictor of spoken language performance, a pattern similar to that of the CTH who also scored predominantly above the 50th percentile.

Why There May Not Be a Relationship for Children Above the 50th Percentile

A divergence emerges between the breadth of vocabulary knowledge some children reveal during norm-referenced vocabulary testing and the more modest vocabulary they actually produce in conversations. In terms of lexical-semantic knowledge, this might be posited as a child having two distinct facets of their vocabularies: a theoretical vocabulary composed of words they can understand or produce in high-support situations and a practical vocabulary composed of words they understand and produce in situations with little support and high cognitive load such as conversations. Furthermore, producing a word during conversations requires a deeper level of word knowledge than what is required to identify it on a norm-referenced test (Anderson & Freebody, 1981; Nation, 2001; Webb, 2013). When conversing, the situation is less predictable than in a testing context, and the child must process multiple words and syntactic structures produced by the conversational partner(s) and respond with multiword, syntactically precise utterances. As a listener in a conversation, a child must understand not only a single word they hear generally without the visual support so common in norm-referenced testing but also many other words in the sentence(s) as well as the syntax used. The cognitive and linguistic demands on a child as a listener during conversations are reduced in norm-referenced testing, and receptive vocabulary testing also provides the opportunity to guess a correct answer. As a speaker, they must do more than find a one-word label for a picture; they must determine what words they need to use to adequately convey their thoughts or feelings and produce them in the proper order. Expressive vocabulary tests do limit the ability to guess; however, the slower pace and the need to only focus on one word at a time make providing a word during these tests considerably less of a challenge than word finding during conversations. The demands of listening and engaging in real time may leave fewer cognitive resources for word retrieval, especially if articulation is also more difficult (Ambrose et al., 2014). Such children may fall back on using a smaller word set, that is, a "default" vocabulary, in order to communicate.

If results on norm-referenced tests and language samples are in agreement (e.g., a child performs well or poorly on both), the natural interpretation is that both assessments are tapping into the construct of lexical-semantic knowledge. One could predict that those with

greater vocabulary knowledge will both score higher on norm-referenced tests and produce a greater variety of words when speaking and that children with less vocabulary knowledge will perform more modestly on both tasks. If this was the case, it seems an unnecessary waste of time and effort to perform both types of assessments, and it would be reasonable to use the norm-referenced measure quicker to administer and score.

Our results and the results of others (Ebert & Pham, 2017; Sénéchal et al., 2008; Souza & Cáceres-Assenço, 2021; Ukrainetz & Blomquist, 2002), however, call this diagnostic strategy into question. For children with hearing loss who score below the 50th percentile, norm-referenced vocabulary assessments appear to be a viable means of determining the need for support when performance is in the below-average range. For these children, test scores are predictive of production during spontaneous spoken language, and overall, these children are not producing a lot of variability in spoken language. Therefore, clinicians can be confident that children who score below average on norm-referenced measures would also have low performance in spontaneous spoken language.

Norm-referenced test scores, however, reveal little about the spontaneous spoken language of children who are DHH who score above the 50th percentile, similar to the lack of relationship between the two seen in CTH. For these children, assuming that a strong test performance negates the need for support services may leave many children who struggle in real-world communicative environments without the support they need and may prevent them from reaching their full language potential. Some children who are DHH perform within the average range on norm-referenced assessments of vocabulary, but they may not be able to capitalize on this knowledge in everyday conversations. Conversations increase cognitive demands on a child, and producing a word in that context may require greater depth of lexical-semantic knowledge. As depth is not assessed on most norm-referenced vocabulary assessments, scoring in the "range of normal" on a norm-referenced vocabulary assessment may not be enough to justify a recommendation of no support services: Results of this study indicate that a follow-up language sample should be conducted because it could reveal difficulties not captured on the norm-referenced assessment.

Caution is warranted for the assumption that because a child *can* recognize a word in a highly supported environment, that they therefore *will* recognize that word in a more demanding context, much less also be able to produce it independently in conversations. Whereas these two facets of vocabulary may be virtually identical for some children, they may not be for others. Some children who have a large theoretical vocabulary, an overall strength, may still require support in utilizing this knowledge in everyday situations such as in the classroom.

Limitations and Future Directions

One limitation of this study is that language samples were only collected on one occasion and, thus, cannot know for sure that they are truly representative of the child's speech. Furthermore, individual factors such as personality, fatigue, and interest could impact how much and what a child said. It is also important to consider whether the conversational language sampling elicitation method used here may have impacted the vocabulary children produced; different topics or a more targeted questioning style may have engendered more varied vocabulary. There is evidence that both conversational partner (Hoff, 2010) and situation (Howe et al., 2014) impact word choice. Building on the current project, an investigation into the impact of conversational context on vocabulary production is planned, as well as examining spontaneous language samples of these children at school entry to examine developmental changes in production. Another potential limitation is that we only examined relationships between language samples and norm-referenced assessments of vocabulary designed to test vocabulary breadth rather than depth. Some norm-referenced assessments designed to measure verbal IQ more broadly rather than vocabulary more narrowly may tap into the depth of a child's vocabulary knowledge. For example, the Wechsler Intelligence Scale for Children–Fifth Edition (Wechsler, 2014) asks a child to explain how two words relate to each other as well as the definitions of words, and the Kaufman Assessment Battery for Children–Second Edition (Kaufman & Kaufman, 2004) Riddles subtest has a child name an object after hearing it described. However, these tests are less commonly used in language assessment batteries by speech-language pathologists and, hence, were not included here. Future work will examine the relationship between language produced during naturalistic language samples and measures designed to explore vocabulary depth.

Conclusions

Norm-referenced assessments are a preferred method of evaluation among speech-language pathologists (e.g., Betz et al., 2013; Ogiela & Montzka, 2021). Many factors influence speech-language pathologist test selection, inclusive of test availability, test features, ease of administration, and perceived restrictions on tests that are likely to be "accepted" by a state- or district-level education system (Ogiela & Montzka, 2021). In the United States, children who receive special education services in public school from a related-services professional such as a speech-language pathologist

must qualify for services on the basis of a documented deficit (e.g., a hearing loss or language delay) that also has an adverse educational impact (Individuals with Disabilities Education Act, 2004). Thus, children who are DHH do not qualify for support services on the basis of a hearing loss diagnosis alone; norm-referenced test scores and academic performance are used to determine a child's support needs. Data from the present study suggest that a single-word vocabulary, norm-referenced measure that indicates a child has a performance above the 50th percentile may not be sufficient to capture nuanced difficulties with conversational vocabulary. This finding aligns with other studies finding that, for children who are DHH, performance in the "range of normal" on a norm-referenced test may not capture ways in which their language performance differs from that of peers with typical hearing (e.g., Lowenstein & Nittrouer, 2021; Lund, 2020). Clinicians should consider adding language sampling to their assessment battery if a child's vocabulary score is above the 50th percentile.

Author Contributions

Leo Evans: Formal analysis, Writing – original draft. **Emily A. Lund:** Conceptualization, Data curation, Writing – review & editing. **Krystal L. Werfel:** Conceptualization, Data curation, Writing – review & editing.

Data Availability Statement

The data sets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

Research reported in this article was supported by National Institute on Deafness and Other Communication Disorders Grants R03DC014535 (principal investigator [PI]: Krystal L. Werfel) and R01DC017173 (PIs: Krystal L. Werfel and Emily A. Lund). Study data were collected and managed using REDCap electronic data capture tools hosted at the University of South Carolina. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Ambrose, S. E., Unflat Berry, L. M., Walker, E. A., Harrison, M., Oleson, J., & Moeller, M. P. (2014). Speech sound production in 2-year-olds who are hard of hearing. *American Journal of Speech-Language Pathology, 23*(2), 91–104. https://doi.org/10.1044/2014_ajslp-13-0039

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Reading comprehension and education* (pp. 77–117). International Reading Association.

Armstrong, R., Scott, J. G., Whitehouse, A. J., Copland, D. A., Mcmahon, K. L., & Arnott, W. (2017). Late talkers and later language outcomes: Predicting the different language trajectories. *International Journal of Speech-Language Pathology, 19*(3), 237–250. https://doi.org/10.1080/17549507.2017.1296191

Betz, S. K., Eickhoff, J. K., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*(2), 133–146. https://doi.org/10.1044/0161-1461(2012/12-0093)

Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test–Fourth Edition*. Academic Therapy Publications.

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 54*(6), 1609–1627. https://doi.org/10.1044/1092-4388(2011/10-0247)

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349–358. https://doi.org/10.1044/1058-0360(2003/080)

de Villiers, J., & Johnson, V. (2007). Implications of new vocabulary assessments for minority children. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 157–181). Guilford Press.

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test –Fifth Edition* [Measurement instrument]. NCS Pearson.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. Pearson.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research, 39*(3), 643–654. https://doi.org/10.1044/jshr.3903.643

Ebert, K. D., & Pham, G. (2017). Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language, Speech, and Hearing Services in Schools, 48*(1), 42–55. https://doi.org/10.1044/2016_lshss-16-0007

Ebert, K. D., & Scott, C. M. (2014). Relationships between narrative language samples and norm-referenced test scores in language assessments of school-age children. *Language, Speech, and Hearing Services in Schools, 45*(4), 337–350. https://doi.org/10.1044/2014_lshss-14-0034

Ehrler, D., & McGhee, R. (2008). *Primary Test of Nonverbal Intelligence*. Pro-Ed.

Evans, J. L. (1996). SLI subgroups: Interaction between discourse constraints and morphosyntactic deficits. *Journal of Speech and Hearing Research, 39*(3), 655–660. https://doi.org/10.1044/jshr.3903.655

Forrester, M. A., & Cherington, S. M. (2009). The development of other-related conversational skills: A case study of conversational repair during the early years. *First Language, 29*(2), 166–191. https://doi.org/10.1177/0142723708094452

Geers, A. E., Nicholas, J., Tobey, E., & Davidson, L. (2016). Persistent language delay versus late language emergence in children with early cochlear implantation. *Journal of Speech,*

Language, and Hearing Research, 59(1), 155–170. https://doi.org/10.1044/2015_jslhr-h-14-0173

Hadley, P. A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools, 29*(3), 132–147. https://doi.org/10.1044/0161-1461.2903.132

Hoff, E. (2010). Context effects on young children's language use: The influence of conversational setting and partner. *First Language, 30*(3–4), 461–472. https://doi.org/10.1177/0142723710370525

Howe, N., Abuhatoum, S., & Chang-Kredl, S. (2014). "Everything's upside down. We'll call it upside down valley!": Siblings' creative play themes, object use, and language during pretend play. *Early Education and Development, 25*(3), 381–398. https://doi.org/10.1080/10409289.2013.773254

Individuals with Disabilities Education Act of 2004, Pub. L. No. 108-446, 118 Stat. 2647. (2004).

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children–Second Edition (KABC-II)*. AGS.

Klee, T. (1985). Clinical language sampling: Analysing the analyses. *Child Language Teaching and Therapy, 1*(2), 182–198. https://doi.org/10.1177/026565908500100206

Law, F., & Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development, 11*(4), 331–355. https://doi.org/10.1080/15475441.2014.961066

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69–92. https://doi.org/10.1017/s0142716410000299

Lowenstein, J. H., & Nittrouer, S. (2021). The devil in the details can be hard to spot: Malapropisms and children with hearing loss. *Language, Speech, and Hearing Services in Schools, 52*(1), 335–353. https://doi.org/10.1044/2020_lshss-20-00033

Lund, E. (2020). The relation between vocabulary knowledge and phonological awareness in children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 63*(7), 2386–2402. https://doi.org/10.1044/2020_jslhr-19-00259

Lund, E., & Douglas, W. M. (2016). Teaching vocabulary to preschool children with hearing loss. *Exceptional Children, 83*(1), 26–41. https://doi.org/10.1177/0014402916651848

Martin, N. A., & Brownell, R. (2011). *Receptive One-Word Picture Vocabulary Test*. Pro-Ed.

Miller, J. F., & Iglesias, A. (2012). *Systematic Analysis of Language Transcripts* (SALT) [Computer software]. SALT Software.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nikolaus, M., & Fourtassi, A. (2023). Communicative feedback in language acquisition. *New Ideas in Psychology, 68,* Article 100985. https://doi.org/10.1016/j.newideapsych.2022.100985

Nowland, K. (2009). *Kentucky eligibility guidelines for students with speech or language impairment–revised*. Kentucky Department of Education.

Ogiela, D. A., & Montzka, J. L. (2021). Norm-referenced language test selection practices for elementary school children with suspected developmental language disorder. *Language, Speech, and Hearing Services in Schools, 52*(1), 288–303. https://doi.org/10.1044/2020_lshss-19-00067

Ramos, M. N., Collins, P., & Peña, E. D. (2022). Sharpening our tools: A systematic review to identify diagnostically accurate language sample measures. *Journal of Speech, Language, and Hearing Research, 65*(10), 3890–3907. https://doi.org/10.1044/2022_jslhr-22-00121

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 49*(1), 74–87. https://doi.org/10.1044/1092-4388(2006/006)

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education and Development, 19*(1), 27–44. https://doi.org/10.1080/10409280701838710

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*(3), 246–259. https://doi.org/10.1044/1058-0360(2007/029)

Souza, M. S. D. L., & Cáceres-Assenço, A. M. (2021). O vocabulário e as habilidades narrativas se correlacionam em pré-escolares com desenvolvimento típico de linguagem? [Do vocabulary and narrative skills correlate in preschoolers with typical language development?]. *Sociedade Brasileira de Fonoaudiologia, 33*(6). https://doi.org/10.1590/2317-1782/20202020169

Spaulding, T. J. (2012). Comparison of severity ratings on norm-referenced tests for children with specific language impairment. *Journal of Communication Disorders, 45*(1), 59–68. https://doi.org/10.1016/j.jcomdis.2011.11.003

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)

Stockman, I. J. (1999). Semantic development of African American children. In O. L. Taylor & L. B. Leonard (Eds.), *Language acquisition across North America: Cross-cultural and cross-linguistic perspectives* (pp. 61–108). Singular.

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test–III: An illusion of unbiased assessment? *Language, Speech, and Hearing Services in Schools, 31*(4), 340–353. https://doi.org/10.1044/0161-1461.3104.340

Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear and Hearing, 36*(Suppl. 1), 76S–91S. https://doi.org/10.1097/aud.0000000000000219

Ukrainetz, T. A., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy, 18*(1), 59–78. https://doi.org/10.1191/0265659002ct227oa

Walker, E. A., Holte, L., McCreery, R. W., Spratford, M., Page, T., & Moeller, M. P. (2015). The influence of hearing aid use on outcomes of children with mild hearing loss. *Journal of Speech, Language, and Hearing Research, 58*(5), 1611–1625. https://doi.org/10.1044/2015_jslhr-h-15-0043

Walker, E. A., Ward, C., Oleson, J. J., Sapp, C., McCreery, R., Tomblin, J. B., & Moeller, M. P. (2022). Aided hearing moderates the academic outcomes of children with mild to severe hearing loss. *Journal of Early Hearing Detection and Intervention, 41*(4), 775–789. https://doi.org/10.1097/aud.0000000000000823

Wallace, G., & Hammill, D. D. (2002). *Comprehensive Receptive and Expressive Vocabulary Test*. Pro-Ed.

Washington, J., & Craig, H. (1992). Performances of low-income African American preschool and kindergarten children on the Peabody Picture Vocabulary Test–Revised. *Language, Speech, and Hearing Services in Schools, 23*(4), 329–333. https://doi.org/10.1044/0161-1461.2304.329

Webb, S. (2013). Depth of vocabulary knowledge. In C. Chappelle (Ed.), *Encyclopedia of applied linguistics* (pp. 1656–1663). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1325

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children–Fifth Edition*. Pearson.

Werfel, K. L. (2017). Emergent literacy skills in preschool children with hearing loss who use spoken language: Initial findings from the Early Language and Literacy Acquisition (ELLA) study. *Language, Speech, and Hearing Services in Schools, 48*(4), 249–259. https://doi.org/10.1044/2017_lshss-17-0023

Werfel, K. L. (2018). Morphosyntax production of preschool children with hearing loss: An evaluation of the extended optional infinitive and surface accounts. *Journal of Speech, Language, and Hearing Research, 61*(9), 2313–2324. https://doi.org/10.1044/2018_jslhr-l-17-0406

Werfel, K. L., & Douglas, M. (2017). Are we slipping them through the cracks? The insufficiency of norm-referenced assessments for identifying language weaknesses in children with hearing loss. *Perspectives of the ASHA Special Interest Groups, 2*(9), 43–53. https://doi.org/10.1044/persp2.sig9.43

Werfel, K. L., Reynolds, G., & Fitton, L. (2022). Oral language acquisition in preschool children who are deaf and hard-of-hearing. *Journal of Deaf Studies and Deaf Education, 27*(2), 166–178. https://doi.org/10.1093/deafed/enab043

Williams, K. T. (2019). *Expressive Vocabulary Test–Third Edition* [Measurement instrument]. NCS Pearson.