Tutorial

Using Developmental Norms for Speech Sounds as a Means of Determining Treatment Eligibility in Schools

Holly L. Storkel[a]

Purpose: For a child to receive treatment of a speech sound disorder in public schools, the child must demonstrate evidence of an exceptionality in producing speech sounds. One method advocated by some state or local guidelines is to use developmental norms for speech sounds to define impaired speech. However, current practices, as codified in state or local guidelines, may not be encouraging optimal use of this data source. The purpose of this article is to outline best practices in using developmental norms to determine eligibility for speech treatment in school settings.
Method: Three commonly used sets of developmental norms (Sander, 1972; Shriberg, 1993; Smit, Hand, Freilinger, Bernthal, & Bird, 1990) are reviewed to generate best practices in the use of developmental norms to establish eligibility for speech treatment in schools. Three clinical scenarios then are used to illustrate implementation of these best practices.
Results: The review of the normative studies indicates that a strict age cutoff used in isolation is counter to the intended use of developmental norms, representing a conceptualization of normal development that is too narrow. Best practice entails using a richer representation of development, specifically reflecting the range and variability inherent in development. Moreover, diagnosing the presence of a speech sound disorder requires more than just a single measure.
Conclusion: Clinicians may need to advocate for change in state or local guidelines to better align these guidelines with best practices in using speech sound norms to determine eligibility for services in schools.

For a student to be eligible for speech treatment in U.S. schools, the assessment team needs to demonstrate that the student has "impaired articulation…that adversely affects a child's educational performance" (Individuals With Disabilities Education Act [IDEA], 2004 (c)(11)). Note that these guidelines focus on articulation, but for the purposes of this article, we assume that articulation encompasses motor and phonological aspects that contribute to speech sound disorders. One method advocated by some state or local guidelines is to use developmental norms for speech sounds to define impaired speech. In some cases, developmental norms are one option among several for determining whether articulation is impaired. For example, Wisconsin indicates that one of four possible options for documenting "delayed speech or sound production" is "consistent speech sound errors when 90% of typically developing children produce the sound correctly" (Wisconsin Department of Public Instruction, 2009). Other guidelines endorse developmental norms as the only option for documenting impaired speech. For example, one school district states that a child can be identified as having a speech impairment if the "student demonstrates speech sound errors outside of the developmental guidelines" (Glencoe District, 2011, p. 3) and then the clinician is directed to appendices containing phonological process and sound norms. Developmental norms for phonological error patterns or speech sounds can be a useful tool in determining eligibility, but some guidelines seem to overemphasize the use of norms, particularly single age cutoffs, and, in some cases, even advocate for use that is inappropriate. In this article, I provide a refresher on developmental norms and consider how to best use developmental norms in determining whether a speech sound disorder is present. Of course, determining eligibility for speech-language intervention in schools encompasses more than just detecting the presence of a speech sound disorder. However, for this article, I focus exclusively on this step

[a]Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas, Lawrence

Correspondence to Holly L. Storkel: hstorkel@ku.edu

Editor: Brenda Beverly

Received July 17, 2018
Revision received November 21, 2018
Accepted December 11, 2018

https://doi.org/10.1044/2018_PERS-SIG1-2018-0014

Publisher Note: This article is part of the Forum: Speech Sound Disorders in Schools.

Disclosures

Financial: Holly L. Storkel has no relevant financial interests to disclose.
Nonfinancial: Holly L. Storkel has no relevant nonfinancial interests to disclose.

and how developmental norms should contribute to this step. Other articles in the forum address the issue of determining educational impact.

Developmental Norms

Several commonly used sources for developmental norms provide guidelines or important considerations for using normative data in clinical decision making. Many of these foundational recommendations have been lost as states and districts have developed guidelines for clinical decision making. A look back at the original recommendations is informative in reconsidering what constitutes best practice.

The norms created by Sander (1972) are currently available in the American Speech-Language-Hearing Association Practice Portal (American Speech-Language-Hearing Association, n.d.). Sander provided a reanalysis of normative data from Wellman, Case, Mengert, and Bradbury (1931) and Templin (1957). Specifically, Sander used Wellman and colleagues' data from 2-year-old children and Templin's data from 480 three- to eight-year-old children to determine two ages of normative sound production. Specifically, Sander defined the average age of (a) *customary production* as the earliest age when "the combined test average at the various word positions exceeds 50% correct production" (p. 61) and (b) *upper age limits* as the earliest age when "the combined test average reaches 90% correct production" (p. 61). Sander then created a chart graphing customary to upper age limit production for each sound to illustrate the variability in and protracted nature of development of each sound. Sander advocated for using this chart to avoid "the pitfalls of attaching a single acquisition age to each consonant" (p. 61) so that one can understand that "normality in articulatory development encompasses an impressively broad age range" (p. 61).

Relative to Sander's (1972) argument about "the pitfalls of attaching a single acquisition age to each consonant" (p. 61), it is important to note that the diagnostic accuracy for different acquisition criteria (e.g., 50% vs. 75% vs. 90%) has not been determined. Thus, the impact of selecting different cutoffs for determining an appropriate age of acquisition is unknown. Any selected cutoff is somewhat arbitrary, similar to selecting a cutoff score on a standardized test without considering the diagnostic accuracy of that cutoff score. Plante and Vance (1994) make the argument that a set standard cutoff score on a standardized test, without reference to diagnostic accuracy for that specific test, will misidentify children with language impairments. The same argument can be applied to normative data for speech sounds: The most diagnostically accurate cutoff for any set of norms is unknown. For this reason, it may be helpful to compare a child's production of each speech sound to a range of expected ages, in the case of the data from Sander, from customary to upper age limits, rather than reducing normal acquisition to a single age of production.

Smit et al. (1990) collected single-word picture-naming data from 997 children of ages 3;0–9;0 (years;months) in Iowa and Nebraska. Each phoneme was targeted in initial and final positions in a variable number of words (two to four). Note that children who were receiving treatment for speech were "included up to the approximate proportion they represented in the population at their age level" (p. 780). Smit and colleagues then defined recommended ages of acquisition that were usually the earliest age when 90% of responses were acceptable. This criterion was chosen because it was the criterion used by both states (Iowa and Nebraska) that commissioned the study. Although this departs from Sander's (1972) recommendation to use more than a single age to represent normal development, Smit and colleagues deemed the criteria appropriate because "states' guidelines also paid appropriate attention to other factors such as total number of errors and overall intelligibility" (p. 793). In this way, Smit and colleagues encouraged the use of multiple measures beyond just a single age of acquisition for sounds to establish eligibility. Ultimately, both authors encourage one to avoid reducing normal development to a single age (as in Sander, 1972) or a single measure (as in Smit et al., 1990). In addition, Smit and colleagues indicated that these norms were only appropriate for individual sound errors. They state, "Certainly, if a child's error on one sound is repeated for other, similar sounds, and if those errors reflect a phonologic pattern or process, then the pattern or process needs to be evaluated against norms for the process rather than norms for individual phoneme targets" (p. 796). This again reinforces that the approach to determining what is normal and what is not normal needs to reflect a variety of measures with normative data for target speech sounds being just one piece of the diagnostic puzzle. That is, guidelines that suggest that a single age of acquisition for each sound is the only or main criterion for determining eligibility for speech services fail to consider the appropriateness of that data source for each child, calling into question the validity of the assessment and resulting clinical decision.

Returning to the issue of diagnostic accuracy in developmental norms, Smit and colleagues (1990) did provide some validation of the 90% criteria as one possible diagnostic cutoff. Specifically, they noted that performance tended to plateau after the 90% criterion had been reached. Thus, if a child had not acquired the sound by the age when 90% of responses were acceptable, then it seemed unlikely that the child would learn that sound without intervention. This fits Sander's (1972) labeling of the 90% criterion as the upper age limits of mastery. Moreover, Smit and colleagues further cautioned against waiting periods, indicating that the data provided "little support for service guidelines that require a child to be 1 year older than the age level at which a 90% criterion is reached before intervention will be provided" (p. 797). Here again, the thinking is that change is unlikely to occur during a waiting period after the upper age limit has been reached. Thus, any guidelines for eligibility that rely on a criterion 1 year beyond the 90% criterion are not consistent with the available evidence.

Lastly, Smit and colleagues (1990) provided detailed and specific recommendations for interpreting the age of

acquisition of /s z/. Although their recommended age of acquisition for /s z/ is quite late (i.e., 7;0–9;0), they once again suggest that a nuanced approach to acquisition of these sounds is key. Specifically, they note that the majority of errors for /s z/ after the age of 6;0 are dental distortions. Thus, a child who continues to make substitution errors after the age of 6;0 would not be showing a typical developmental pattern in terms of type of error. In addition, they note that other distortions of /s/ and /z/, especially lateralized productions or those with "damaging social consequences," should be considered for treatment at or before the age of 7;0, indicating early intervention "even for preschoolers" might be appropriate (p. 796). Once again, Smit and colleagues reinforce that it is not enough to know which sounds are in error and at what ages those sounds are typically in error but the kinds of errors being made deserve consideration as well. As a result, eligibility guidelines should encourage the use of multiple measures of development.

Shriberg (1993) took a slightly different approach, examining the conversational speech of 64 three- to six-year-old children with speech sound disorders. In examining the percent correct for different speech sounds, he noticed natural breaks that grouped the 24 consonants examined into three sets of eight. Early eight consonants included /m n w j h b p d/, which averaged 75% correct productions or better. The mid eight consonants were 25%–75% correct, on average, and included /ŋ t k g f v ʃ ʤ/. The late eight consonants averaged less than 25% accurate productions and consisted of /θ ð s z ʃ ʒ l r/. This grouping of sounds into broad categories of age of acquisition addresses the issue raised by Sander (1972) in a slightly different way by not assigning a specific age of acquisition at all but rather noting broad patterns in when groups of sounds are acquired. Shriberg cross-referenced these groupings with normative studies of speech sound acquisition from Smit and colleagues (1990) and Sander, among others. In general, there was relatively good agreement between these different sources in categorizing sounds as early, mid, or late eight. It was noted that disagreements were likely due to methodological differences across studies, such as different sampling techniques (e.g., conversational speech vs. single-word picture naming) and different criteria for defining mastery. An important insight from this comparison of Shriberg's data from children with speech sound disorders and prior data on children with predominately typical development is that the sequence of consonant mastery is relatively similar between children with speech sound disorders and typically developing children. Shriberg then combined these normative data about speech sound acquisition with other sources of information, including word shapes and types of errors, to create his Speech Disorders Classification System. As with Smit and colleagues, Shriberg's Speech Disorders Classification System indicates that developmental norms for speech sounds are only one component that should be integrated with other components in identifying a speech sound disorder.

Best Practice Guidelines

Taken together, reexamination of the original normative studies suggests the following two best practices for using age of acquisition of individual speech sounds in determining eligibility for services. First, age of acquisition of each speech sound should not be considered as just a single age or a sharp cut-point because there is variability across studies in the age of acquisition for each sound and there is no validation of diagnostic accuracy for any given cut-point. Therefore, best practice entails consideration of the range of age of acquisition. This can be accomplished by using the chart created by Sander (1972), showing the range of customary to upper age limit definitions for acquisition for each individual sound. Alternatively, if using the norms from Smit and colleagues (1990), the 75% (see Tables 5 and 6) and 90% (see Table 7) ages could be presented for each speech sound or the exact percentage of acceptable responses at the child's age (see Table 4) could be reported for each speech sound. Alternatively, the general categories identified by Shriberg (1993) could be used to summarize which groups of sounds are accurate versus inaccurate. Second, normative data need to be integrated with other measures. Speech sound acquisition is too complex to be reduced to a single data point to represent normal development. Smit and colleagues mention examination of total number of errors and intelligibility and also implied that error patterns (e.g., phonological processes) and types of errors (e.g., substitutions vs. distortions) needed to be factored in as well. Shriberg also indicates the importance of error analysis and word shapes.

Application of Developmental Norms

Next, I apply these concepts to the three scenarios described in the introduction to this issue, using a reduced assessment battery to illustrate how even just a few additional measures can be integrated with age of acquisition data. That is, these illustrations are not intended to illustrate a comprehensive assessment battery to determine eligibility. Recall that the three scenarios were all 6-year-old children. I assume that they are male children due to the gender disparity in the prevalence of speech sound disorders (Shriberg, Tomblin, & McSweeny, 1999). One child made multiple errors on late-acquired sounds (/ŋ s z r/). One child made an error on a single sound (/r/). One child made distortion errors on a group of sounds (lateralized /s z ʃ ʧ ʤ/). For each case, I used a range of normative data from Smit and colleagues (1990) rather than a single cutoff for age of acquisition, in keeping with the best practice guidelines outlined above. For an additional measure, I chose to simulate performance on a standardized test of articulation. Although many options exist for measures to combine with developmental norms, my assumption was that many speech-language pathologists use standardized tests to obtain a speech sample. Moreover, several of the recently published standardized tests of articulation now report

diagnostic accuracy. At least three recently published tests and one older test report adequate diagnostic accuracy based on criteria outlined in Plante and Vance (1994): Arizona Articulation and Phonology Scale–Fourth Edition (Fudala & Stegall, 2017), Clinical Assessment of Articulation and Phonology–Second Edition (Secord & Donohue, 2014), Diagnostic Evaluation of Articulation and Phonology (Dodd, Hua, Crosbie, Holm, & Ozanne, 2006), and Goldman-Fristoe Test of Articulation–Third Edition (Goldman & Fristoe, 2015). Thus, a score on one of these tests with documented acceptable diagnostic accuracy would be an important piece of normative information in determining eligibility, indicating how the total number of errors compares to same-age peers. In addition, the speech sample obtained for a standardized test can be used to identify sounds in error for comparison to developmental norms as well as additional error patterns (e.g., type of error) to complement developmental norms. For each case, I entered the described pattern on the Arizona Articulation Scale–Fourth Edition (Fudala & Stegall, 2017) and computed the raw score, standard score, and percentile rank for boys (see Figures 1a, 2, and 3). For the first child, I also completed the Arizona Phonology Scale–Fourth Edition (Fudala & Stegall, 2017; see Figure 1a) and assumed that the child exhibited fairly typical substitution or deletion patterns for each sound in error.

Scenario 1

The simulation for the first child who showed multiple errors on late-acquired sounds (/ŋ s z r/) is shown in Figures 1a and 1b. As shown in the left panel of Figure 1a, errors on these four sounds resulted in 18 total simulated errors on the Arizona Articulation Scale–Fourth Edition (Fudala & Stegall, 2017). This simulated performance yielded a standard score of 62, which corresponded to the first percentile. This score is below the recommended cut score of 78, which showed good sensitivity and specificity based on information reported in the test manual. This score indicates that the child has too many errors compared to his same-age peers. Analysis of phonological error patterns is shown in the right panel of Figure 1a. A phonological error pattern analysis may be appropriate when multiple sounds are in error, serving to identify patterns that are common across target sounds. However, Figure 1a shows that phonological error pattern use was relatively low with each individual target showing a unique pattern, apart from stopping of both /s/ and /z/. Thus, ultimately, a phonological error pattern analysis does not yield many new insights for this child, but the Arizona Phonology Scale–Fourth Edition (Fudala & Stegall, 2017) does provide normative data for phonological error patterns. These data indicate that at least 50% of children would have eliminated these phonological error patterns at an age younger than 6;0. However, the age when 90% of children would have eliminated these phonological error patterns is beyond the age of 6;0. Thus, the picture that is emerging is that this child produces too many errors (standardized test score) but that the error patterns are consistent with upper age limit expectations (90% criterion).

Figure 1b shows the relevant normative data from Smit and colleagues (1990). Here, the value of presenting detailed normative data is apparent. If one were to examine only the 90% age of acquisition, concern for this child might be minimized because the 90% age of acquisition for the sounds in error is well beyond the child's chronological age of 6;0. This seems at odds with the standard score and percentile from the Arizona Articulation Scale–Fourth Edition (Fudala & Stegall, 2017). However, a different picture emerges when more detailed normative data are considered. Specifically, almost all the sounds in error (except for several clusters) show a 75% age of acquisition at the age of 6;0 or earlier. In addition, the percentage of acceptable responses by boys at the age of 6;0 shows that there are a high number of acceptable responses across all these targets at the age of 6;0. This more detailed normative picture converges with the standard score and percentile from the Arizona Articulation Scale–Fourth Edition. That is, the number of errors exhibited by this child appears to be greater than his peers (by standard score on the Arizona Articulation and Phonology Scale–Fourth Edition), even though the specific sounds in error are not unusual (by normative data). We can integrate these observations to conclude that a child this age should be demonstrating some acceptable productions for at least some of these sounds. Moreover, a closer reading of Smit and colleagues reveals that stopping of /s z/ is a different error pattern than what was observed for same-age peers, who were noted to produce dentalized distortions for /s z/, rather than substitutions (i.e., stopping). Taken together, standardized test scores and detailed normative data converge on the conclusion that this child presents with an exceptionality in speech sound development and may be eligible for services, pending evaluation of the educational impact of this exceptionality. It is important to keep in mind that, although the minimal data applied to this case converge on the conclusion that a speech sound disorder is present, more information is needed for a comprehensive assessment and treatment planning. Given the multiple errors present in this case, assessment of both accuracy (e.g., percent consonants correct) and intelligibility (e.g., percent intelligibility) in connected speech would be an important component of a comprehensive assessment to better understand the impact of multiple speech sound errors on communication.

Scenario 2

Turning to the second child, who showed errors only on /r/, the upper left panel of Figure 2 shows that this child's error pattern would yield nine total errors on the Arizona Articulation Scale–Fourth Edition (Fudala & Stegall, 2017), corresponding to a standard score of 85 and a percentile rank of 16. This score is above the recommended cutoff, according to the test manual, and suggests that the child's number of errors is like that of his same-age peers. Detailed normative data from Smit and colleagues

Figure 1. (a) Simulated scores and patterns on the Arizona Articulation Scale–Fourth Edition (left) and Arizona Phonology Scale–Fourth Edition (right) for the first scenario: a 6-year-old boy who showed multiple errors on late-acquired sounds /ŋ s z r/. (b) Normative data (Smit et al., 1990) for singletons (left) and clusters (right) for the first scenario: a 6-year-old boy who showed multiple errors on late-acquired sounds /ŋ s z r/.

(a)

Simulated Score on the Arizona Articulation Scale-4

	# of Errors	Type of Error
ŋ	1 of 1	Fronting: /ŋ/ → /n/
r	3 of 3 word initial 6 of 6 word final	Gliding: /r/ → /w/ Vowelization: /r/ → /Ø/
s	2 of 2 singletons 4 of 4 clusters	Stopping: /s/ → /t/ Cluster reduction: /CC/ → /C/
z	2 of 2	Stopping: /z/ → /d/
	Total Errors: 18 Total Error Value: 21.5 Raw Score: 78.5	Standard Score = 62 Percentile Rank = 1

Simulated Score on the Arizona Phonology Scale-4

	# of Errors	# of Opps	%	50% No Errors[1]	90% No Errors[1]
Cluster simplification	4	11	36%	4;0-4;5	7;0-7;5
Fronting	1	24	4%	3;6-3;11	6;6-6;11
Stopping	5	22	23%	4;0-4;5	7;0-7;5
Gliding	6	12	50%	4;6-4;11	7;0-7;5
Vowelization	8	9	89%	3;0-3;5	7;0-7;5

[1]Age at which 50% or 90% of the normative sample produced zero errors for this phonological error pattern.

(b)

Smit et al. (1990) Normative Data

	75% AoA[1]	90% AoA[2]	% Male 6;0[3]
ŋ	</= 3;0	7;0-9;0	75%
r	5;6	8;0	76% initial 82% final
s	5;0	7;0-9;0	79%[4]
z	6;0	7;0-9;0	75%[4]

[1]Table 5, [2]Table 7, [3]Table 4, [4]"By age 6;0, virtually all error types were dental." (p. 796)

	75% AoA[1]	90% AoA[2]	% Male 6;0[3]
pr	5;6	8;0	75%
br	6;0	8;0	76%
tr	5;6	8;0	76%
dr	5;0	8;0	78%
kr	5;6	8;0	78%
gr	5;6	8;0	75%
fr	5;6	8;0	75%
θr	7;0	9;0	60%
sp	5;0	7;0-9;0	81%
st	5;0	7;0-9;0	81%
sk	6;0	7;0-9;0	79%
sm	7;0	7;0-9;0	71%
sn	5;0	7;0-9;0	76%
sw	6;0	7;0-9;0	78%
sl	7;0	7;0-9;0	72%

[1]Table 5, [2]Table 7, [3]Table 4

(1990), shown in the right panel of Figure 2, reveal a similar picture. The 75% age of acquisition is near the child's chronological age for most sounds in error, but the 90% age of acquisition is beyond the child's chronological age for all sounds in error. Likewise, error patterns observed by Smit (1993a, 1993b), denoted in the lower left panel of Figure 2, show that this child's error patterns are like those of other children. Once again, detailed normative data converge with standardized test scores to suggest that this child's speech sound development falls in the low normal range. Based on this brief snapshot of the child's speech, it is difficult to make a compelling case that an exceptionality in speech sound development is present based solely on the standardized test and developmental norms. However, the information presented here is incomplete and not representative of a full and detailed assessment. It is possible that additional information about the child's speech sound development might uncover areas of concern. For example, assessing stimulability for /r/ would be an important step in evaluating the likelihood that the child would be able to improve production of /r/ without any additional support. Likewise, as detailed by the other articles in this forum, it is critical to consider the impact of this child's speech sound error in terms of his ability to communicate effectively, his willingness to communicate with peers and adults, and his performance in related areas that rely on

Figure 2. Simulated scores and patterns on the Arizona Articulation Scale–Fourth Edition and normative data (Smit, 1993a, 1993b; Smit et al., 1990) for the second scenario: a 6-year-old boy who misarticulated /r/ in singletons and clusters.

Simulated Score on the Arizona Articulation Scale-4

	# of Errors	Type of Error
r	3 of 3 word initial 6 of 6 word final	Gliding: /r/ → /w/ Vowelization: /r/ → /Ø/
	Total Errors: 9 Total Error Value: 10 Raw Score: 90	Standard Score = 85 Percentile Rank = 16

Smit (1993a, 1993b) Error Patterns for Age 6;0

Sound	Error Pattern
/r/initial /r/ final	/r/ → /w/ /r/ → /Ø/ (rounded vowel)
Cr	Cr → Cw

Smit et al. (1990) Normative Data

	75% AoA[1]	90% AoA[2]	% Male 6;0[3]
r	5;6	8;0	76% initial 82% final
pr	5;6	8;0	75%
br	6;0	8;0	76%
tr	5;6	8;0	76%
dr	5;0	8;0	78%
kr	5;6	8;0	78%
gr	5;6	8;0	75%
fr	5;6	8;0	75%
θr	7;0	9;0	60%

[1]Table 5, [2]Table 7, [3]Table 4

phonology (e.g., literacy). Only after investigating these issues will we have a clear picture of the child's strengths and weaknesses to fully evaluate eligibility and need for services as well as appropriate service delivery options.

Scenario 3

Moving on to the third and final child who showed distorted (i.e., lateralized) production of /s z ʃ ʧ ʤ/, the left panel of Figure 3 shows that this child's error pattern would yield 13 productions in error on the Arizona Articulation Scale–Fourth Edition (Fudala & Stegall, 2017). These errors correspond to a standard score of 78 and a percentile rank of 7. This score is at the recommended cutoff score, according to the test manual, indicating that the child's performance is significantly below that of his same-age peers and likely indicative of a speech sound disorder. Detailed normative data are shown in the right panel of Figure 3. As with the other cases, the 75% age of acquisition is at or below the child's chronological age, but the 90% age of acquisition is beyond the child's chronological age. Likewise, there are a large percentage of acceptable responses for 6-year-old boys for each target sound in the normative sample. In some sense, this case looks like the first case with the total number of errors being the cause for concern based on both the test score and the detailed normative data. Unlike the first case, the normative data provide further relevant information. In particular, lateralized productions for the target sounds were noted to be infrequent (0%–5% of the total responses) at the

Figure 3. Simulated scores and patterns on the Arizona Articulation Scale–Fourth Edition and normative data (Smit, 1993a; Smit et al., 1990) for the third scenario: a 6-year-old boy with distorted (i.e., lateralized) production of /s z ʃ ʧ ʤ/.

Simulated Score on the Arizona Articulation Scale-4

	# of Errors	Type of Error
s	6 of 6	Distortion: Lateralized
z	2 of 2	Distortion: Lateralized
ʃ	2 of 2	Distortion: Lateralized
ʧ	2 of 2	Distortion: Lateralized
ʤ	1 of 1	Distortion: Lateralized
	Total Errors: 13 Total Error Value: 13.5 Raw Score: 86.5	Standard Score = 78 Percentile Rank = 7

Smit and Colleagues (1990) Normative Data

	75% AoA[1]	90% AoA[2]	% Male 6;0[3]
s	5;0	7;0-9;0	79%[4]
z	6;0	7;0-9;0	75%[4]
ʃ	5;0	7;0	88%[5]
ʧ	5;0	7;0	89%[5]
ʤ	4;0	7;0	92% initial[5] 87% final[5]

[1]Table 5, [2]Table 7, [3]Table 4, [4]"By age 6;0, virtually all error types were dental.... Lateralizations and post-alveolar distortions were 0-5% of the total responses, depending on the age group." (p. 796) [5]According to Smit (1993b), lateralized distortions were < 5% of the total responses in all age groups for these targets.

age of 6;0 and beyond. In summary, the standardized test score indicates that this child is producing too many errors for his chronological age, and detailed normative data indicate that the type of error (i.e., lateralized) is not typically observed at this age. Thus, an exceptionality exists, and educational impact would need to be evaluated to further establish eligibility and recommendations. Examination of the social consequences of the child's speech sound distortions and the child's self-perception of his communication skills would be particularly important given that the lateralized distortion is not common among typically developing children. It is likely that this child sounds noticeably different from his peers. Consequently, he may receive negative feedback from his peers about his speech (e.g., teasing), which could negatively affect his self-confidence in communicating with others and participating in class.

Best Practice Summary

Developmental norms can be a useful measure in determining the presence or absence of a speech sound disorder. However, current practices, as codified in state or local guidelines, may not be encouraging optimal use of this data source. A strict age cutoff used in isolation does not represent best practice, as outlined in the original published norms. Best practice entails using a richer representation of development, specifically reflecting the range and variability inherent in development. Moreover, diagnosing the presence of a speech sound disorder requires more than just a single measure. The diagnostic accuracy of different cutoffs for developmental norms has not been established. Recent standardized tests of articulation do have acceptable diagnostic accuracy. However, even when diagnostic accuracy meets conventions for acceptability, no single measure is 100% accurate. This means that an accurate diagnosis can only be achieved by combining multiple measures to converge on a diagnosis. Developmental norms are one source of information. When coupled with additional measures that reflect different aspects of speech sound development (e.g., test scores that reflect total number of errors, normative data about error types and patterns), a well-reasoned and likely accurate diagnosis can be achieved. The other articles in this forum provide additional ideas about the different data sources needed to accurately detect the presence of a speech sound disorder and determine the educational impact of a speech sound disorder on a child. The three clinical scenarios, as described in this article and the others in the forum, demonstrate how best practice approaches can yield converging evidence for clinical decision making.

Advocating for Change

State and local guidelines may need to be revised to better reflect current best practice in determining the presence versus absence of a speech sound disorder as well as its impact on educational performance. In general, guidelines should be written to empower the assessment team to apply their clinical decision-making abilities and should avoid overly restrictive eligibility criteria. Thus, one option might be to abandon local guidelines all together and encourage assessment teams to rely on their own expertise and their own interpretation of relevant laws and policies regarding eligibility for services in schools. However, some may argue that laws and policies regarding eligibility are complex and that local guidelines can be important in assisting professionals to translate laws and policies into practice and in creating a shared understanding among team members of how laws and policies should be implemented.

If local guidelines are retained, they should emphasize obtaining a comprehensive set of measures and integrating the results of each measure to better understand and document a specific child's strengths and weaknesses as well as the impact of those strengths and weaknesses on educational performance. In general, single criteria for eligibility should be avoided, and any specific criteria listed in a guideline should be evidence based. Consequently, guidelines that indicate that children must make errors on one or more sounds after the age when 90% of typically developing children produce the sound correctly should be avoided. This type of guideline is (a) too restrictive, requiring that only one source of information be used to determine eligibility (i.e., developmental norms for accuracy), and (b) not evidence based because there is no evidence indicating that these criteria would accurately differentiate children with versus without a speech sound disorder or children whose educational performance is affected by their speech versus children whose educational performance is unaffected by their speech. Checklist approaches are similarly problematic. On the surface, a checklist appears to encourage a multiple-measure approach because several options for qualifying a child for services are offered. However, a checklist is just a selection of several single-criterion options because, ultimately, the speech-language pathologist must check one of the criteria to document a child's eligibility for services. Thus, the checklist approach still reduces eligibility into a binary decision based on a single measure. Choosing one single eligibility criterion from a selection of options is not the same as integrating multiple measures to understand a child's strengths and weaknesses.

At this point, it should be clear that succinctly providing guidance on complex eligibility decisions is not an easy task. One method that seems more promising than single criteria or checklists is a rating scale approach (cf. Colorado Department of Education, 2018). Here, different measures can be included as items to be rated, such as developmental norms, standardized test scores, phonological error patterns, intelligibility, and stimulability, among others. This allows multiple pieces of information to be considered in making the eligibility decision. Each item can then be rated in terms of the level of concern (e.g., none, mild, moderate, severe), which provides a more nuanced interpretation than a binary normal versus disordered decision, as with the single criterion or checklist approaches.

One would still need to consider how to rate developmental norms (as well as other items) within this approach, but presumably one would want to consider the number of sounds a child misarticulates in combination with the normative range of acquisition for each sound. For example, in Clinical Scenarios 1 and 3, both children misarticulated many sounds that had a 75% age of acquisition that was at the child's chronological age or younger and the percentage of acceptable responses at the child's chronological age was high. Therefore, the developmental norm item for these children might be rated as a moderate or severe concern. In contrast, the second clinical scenario showed a child with one sound in error, and that sound had a 75% age of acquisition that was at or near the child's chronological age. For this child, the developmental norm item might be rated as a mild concern or possibly no concern. In creating the ratings, it would be important to think through the different patterns that might be observed so that the different levels of concern for a specific item (e.g., developmental norms) could be clearly defined. It likely would be helpful for groups of speech-language pathologists to work on the rating system together and to do so in the context of clinical cases. That is, each person could bring one example of mild, moderate, and severe cases to help guide the description of the different levels of concern. The cases would not have to be shared with each person but rather could be generically described (e.g., all errors are on sounds with a 75% age of acquisition older than this child's age).

One potential area of concern in using a rating scale approach is that some rating scales seem to require that all items be rated, which is a concern because an assessment should be tailored to the needs of an individual child. Another potential concern is that some rating scales assign points to each level of concern for each item and then the total points are summed to arrive at an overall level of concern or an overall severity rating. This assumes that each item rated is equally important or equally weighted when making the overall diagnostic decision. This assumption requires careful reflection and thought relative to the items being rated because it seems possible that some items could be more influential than others in contributing to an overall diagnosis or severity rating. If this assumption seems unlikely, then the overall diagnosis or severity rating could be decided by the speech-language pathologist or assessment team, rather than being computed by summing individual ratings. In this way, the rating scale would be a way to succinctly summarize findings from multiple measures to support integration and interpretation of these findings to make an eligibility decision.

Conclusions

Current state or local guidelines may not be encouraging optimal use of normative data in determining eligibility for speech treatment in school settings. The literature reviewed suggests that best practices in using developmental norms to establish eligibility for services are as follows: (a) The range and variability inherent in normal development need to be conveyed and considered in clinical decision making, and (b) multiple measures need to be combined to converge on an eligibility decision. State and local guidelines may need to be updated to align with these suggestions, presenting an opportunity to question how guidelines can more successfully support effective clinical decision making.

Acknowledgments

Open access fees were paid for by an award to Holly Storkel from the Friends of the Life Span Institute at the University of Kansas.

References

American Speech-Language-Hearing Association. (n.d.). *Age of customary consonant production*. Retrieved from https://www.asha.org/uploadedFiles/ASHA/Practice_Portal/Clinical_Topics/Late_Language_Emergence/Consonant-Acquisition-Chart.pdf

Colorado Department of Education. (2018). *Communication rating scales: Articulation/phonology*. Retrieved from https://www.cde.state.co.us/cdesped/2018crs_artic_phono

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic Evaluation of Articulation and Phonology*. San Antonio, TX: Pearson. Retrieved from https://www.pearsonclinical.com/language/products/100000295/diagnostic-evaluation-of-articulation-and-phonology-deap.html

Fudala, J. B., & Stegall, S. (2017). *Arizona Articulation and Phonology Scale–Fourth Edition (Arizona-4)*. Torrance, CA: Western Psychological Services. Retrieved from https://www.superduperinc.com/products/view.aspx?pid=TM895#.W0YsC7gnaUk

Glencoe District. (2011). *Speech-language eligibility/dismissal criteria*. Retrieved from http://www.glencoeschools.org/images/district/sleligibility.pdf

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation–Third Edition (GFTA-3)*. Bloomington, MN: Pearson. Retrieved from https://www.pearsonclinical.com/language/products/100001202/goldman-fristoe-test-of-articulation-3-gfta-3.html

Individuals With Disabilities Education Act. (2004). *Section 300.8 child with a disability*. Retrieved from https://sites.ed.gov/idea/regs/b/a/300.8

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15–24. https://doi.org/10.1044/0161-1461.2501.15

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*(1), 55–63. https://doi.org/10.1044/jshd.3701.55

Secord, W. A., & Donohue, J. S. (2014). *Clinical Assessment of Articulation and Phonology–Second Edition (CAAP-2)*. Greenville, SC: Super Duper Publications. Retrieved from https://www.superduperinc.com/products/view.aspx?pid=CAAP12#.W0eah7gnaUk

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36,* 105–140. https://doi.org/10.1044/jshr.3601.105

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and*

Hearing Research, 42, 1461–1481. https://doi.org/10.1044/jslhr.4206.1461

Smit, A. B. (1993a). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research, 36,* 533–547. https://doi.org/10.1044/jshr.3603.533

Smit, A. B. (1993b). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Word-initial consonant clusters. *Journal of Speech and Hearing Research, 36,* 931–947. https://doi.org/10.1044/jshr.3605.931

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779–798. https://doi.org/10.1044/jshd.5504.779

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. Minneapolis, MN: University of Minnesota Press.

Wellman, B., Case, E., Mengert, I., & Bradbury, D. (1931). Speech sounds of young children. *University of Iowa Studies in Child Welfare, 5,* 7–80.

Wisconsin Department of Public Instruction. (2009). *Special education eligibility worksheets and guides*. Retrieved from https://dpi.wi.gov/sped/laws-procedures-bulletins/laws/eligibility