Clinical Linguistics & Phonetics, January–February 2005; 19(1): 53–66

Taylor & Francis healthsciences

Measuring articulatory error consistency in children with developmental apraxia of speech

STACY K. BETZ, & CAROL STOEL-GAMMON

University of Washington, Seattle, WA, USA

(Received 30 October 2003; accepted 27 January 2004)

Abstract

Error inconsistency is often cited as a characteristic of children with speech disorders, particularly developmental apraxia of speech (DAS); however, few researchers operationally define error inconsistency and the definitions that do exist are not standardized across studies. This study proposes three formulas for measuring various aspects of articulatory error consistency: proportion of errors, consistency of error types, and consistency of the most frequently used error type. Each formula is explained using examples of productions from children with DAS and phonological delay. Clinical implications for the use of error consistency to differentially diagnosis DAS and phonological delay are discussed.

Keywords: *Developmental apraxia of speech, articulation, error consistency*

Introduction

Inconsistency of word productions is a hallmark of early child speech. In an oft-cited example from Ferguson and Farwell (1975), a 15-month-old girl produced the word *pen* in ten different ways within a 30-minute period, with pronunciations ranging from [dɛ^{d}] to [hɪn] to [^{m}bõ] to [ba^{h}]. Although none of the ten versions was accurate, each rendition included one or more features of the target form, such as nasality, a labial or alveolar consonant, consonant-vowel-consonant word shape, and/or a stop consonant. As children grow older, their word productions become more consistent; they may not be entirely correct, but they are produced in the same way each time (e.g. *rabbit* is consistently produced as [wæbɪt]). Consistency of production, or lack thereof, has received considerable attention in investigations of children with speech disorders. Inconsistency (often referred to as variability) has been identified as one of the key diagnostic markers of a particular type of disorder: developmental apraxia of speech (Dodd and McCormack, 1995; Davis, Jakielski and Marquardt, 1998; Forrest, 2003). Inconsistency has also served as a source of prognosis for children involved in speech therapy (Forrest, Dinnsen and Elbert, 1997; Tyler, Lewis and Welch, 2003). Therefore, accurate measurement of consistency has implications for both assessment and treatment of children with speech disorders.

Correspondence: Stacy K. Betz, University of Kansas, Child Language Program, Dole Center, 1000 Sunnyside Avenue Room 3031, Lawrence, KS 66045-7555, USA. E-mail: skbetz@ku.edu

ISSN 0269-9206 print/ISSN 1464-5076 online
DOI: 10.1080/02699200512331325791

Even though many studies refer to 'variability' and/or 'inconsistency' of productions, relatively few define the terms and even fewer provide mathematical formulas for computing these aspects of production. Additionally, measures of variability and inconsistency vary from study to study due to the manner in which data were collected. Precise explanations of the manner in which consistency is measured are needed because different types of error consistency can occur, and if lack of consistency is to be used as a diagnostic speech characteristic, it is essential that researchers and clinicians agree on definitions and measures of this phenomenon.

Inconsistency in typical and atypical speech development

At least three types of inconsistency have been identified for typically developing children: (1) inconsistent use of a phoneme based on word position, e.g. /s/ may be produced accurately in word-final position but may be rendered as [t] in word-initial position; thus, /s/ is correct in *bus* and *nice*, but incorrect in *sun* and *soup*; (2) inconsistent use of a phoneme based on the lexical target, e.g. /m/ may be correct in *mommy* but not in *milk* (Ferguson and Farwell, 1975); and (3) inconsistent pronunciations across multiple productions of the same word, as with the child cited in the introduction who produced 10 different forms of the word *pen* within a single data collection session. Given that normal phonological development entails moving from simple, relatively undifferentiated forms (e.g. [da] for *doggie*; [babu] for *bottle*) to adult-like pronunciations, the occurrence of inconsistent productions in early speech is part of the developmental process; however, little is known about the prevalence and persistence of inconsistency in the productions of young typically developing children. Davis and Velleman (2000) state that productions become stable after children attain a vocabulary of 50 words; in contrast, Sosa and Stoel-Gammon (2003) report high degrees of variability well after the first-50-word period.

Two types of explanations for the use of inconsistent productions in children with typical speech and language development have been proposed: (1) incomplete (or 'fuzzy') underlying representations; according to this view, lack of sufficient detail causes the child's productions of a phoneme or a word to vary from one utterance to the next; and (2) inadequate articulatory abilities that prevent the child from producing the correct surface (i.e. spoken) form; according to this view, the underlying representation contains all information needed for correct articulation of a word. Although inconsistency in typically developing child speech is well documented, the underlying causes of the inconsistency remain elusive and the causes may vary across children.

The underlying causes of inconsistency in the speech of children with phonological disorders also have not been determined. Forrest and colleagues (Forrest *et al.*, 1997) argue that a lack of categorical representation of the phoneme is associated with positionally based substitutions for a target consonant, as in the case of /s/ cited above. Interestingly, their research with children with phonological disorders showed that those children who had consistent substitutions for a particular target learned the target sound in treatment and generalized this knowledge to other word positions. Children who had inconsistent substitutions within and across word positions also learned the treated sound in the treated position; however, these children did not generalize the new production to other word positions.

Dodd (1995) states that children with disordered phonology whose speech is characterized by inconsistency have a deficit in 'phonological planning' (i.e. an incomplete or degraded phonological plan) that results in productions in which the 'articulatory parameters' are too broad. Bradford and Dodd (1996) hypothesize that inconsistent speech output may be due to the fact that children with phonological planning deficits may have to create a new plan each time they produce a particular word.

Definitions and measures of consistency

Before the causes of inconsistent productions can be understood, researchers must agree upon definitions and measures of consistency. Although the terms *consistency*, *inconsistency*, and *variability* occur frequently in the literature, published descriptions of ways to measure these phenomena are few and discrepancies across findings abound. Examples of studies providing explicit definitions and/or formulas for computing error consistency include Schumacher, McNeil, Vetter and Yoder (as cited in Hall, Jordan and Robin, 1993), Dodd (1995), Shriberg, Aram and Kwiatkowski (1997), Ingram (2002) and Tyler *et al.* (2003). Each of these studies is described below, followed by a comparison of the definitions, and a discussion of the limitations of these measures.

Schumacher, McNeil, Vetter, and Yoder (as cited in Hall *et al.*, 1993) used measures of consistency and variability in a study comparing the speech of children with DAS and children with functional articulation problems. They defined consistency as the number of erred productions of a given word across multiple productions of the same word and variability as the number of different errors on multiple repetitions of the same word. The authors found that the productions of children with DAS were both consistent and variable, whereas children with functional articulation problems were consistent but not variable in their productions.

Dodd (1995) developed a measure of whole-word inconsistency to classify speech-disordered children into sub-groups; she elicited three productions of a set of 25 words and then examined differences across productions of each word. First, the three trials of each of the 25 target words were labelled as variable or non-variable, with variable production of a given word defined as the use of different productions on at least two of the three trials. Then, each child was classified as displaying either an inconsistent or consistent disorder. Inconsistent disorder was defined as variable production of at least ten of the 25 words. Dodd's goal was not to develop a measure of inconsistency *per se*, but to design a set of procedures that would allow her to categorize different types of speech disorders. Therefore, this measure provides only a nominal level of measurement rather than a continuous value of error consistency.

Shriberg *et al.* (1997) proposed a formula for calculating error consistency in order to describe the speech of children with DAS. This calculation compared multiple productions of the same word occurring during a spontaneous speech sample. Although their formula allows for a combined measure of consistency across multiple words, the formula is described below based on consistency of a single word in order to facilitate comparisons among other definitions of inconsistency. When looking at the productions of one word, error consistency was computed as:

$$(\# \text{ of occurrences of the most frequent error class} - 1)/(\# \text{ of tokens of the word} - 1) \times 100.$$

The most frequent error class was the phoneme deletion or substitution error most common for the target word. For example, if the word *squirrel* occurred three times in the sample as [skɚl], [kɚl], [skɚl], the most frequent error class would be deletion of /w/ and the error consistency of the word *squirrel* would be $(2-1)/(3-1) \times 100 = 50\%$. This measurement of error consistency differs from that used by Dodd (1995) because it is computed based on the occurrences of the most frequent type of error rather than the number of overall occurrences of the target word.

Ingram (2002) introduced a measure of error consistency called 'the proportion of whole-word variation' that compares the number of distinct forms of a given word to the overall number of productions of the same word. Distinct forms for a given word were identified as a difference in production in any of the phonemes of the target word. For example, [ta] and [tat] were considered two distinct forms of the target *cat*. The proportion of whole-word variation was calculated as the number of distinct forms divided by the number of productions of a given word. For example, if the target *cat* was produced three times, as [ta], [ta], and [tat], the proportion of whole-word variation would be 2/3=0.67. Ingram suggested this measure could provide information on a child's underlying phonological representation of a word.

As part of an intervention study of young children with speech and language disorders, Tyler *et al.* (2003) developed the 'error consistency index', a measure of the total number of different sound substitutions/omissions across words and word positions. This index was calculated at the phoneme level, as opposed to the word level used by Dodd (1995), Shriberg *et al.* (1997), and Ingram (2002). To compute the index, the number of different phoneme substitutions/omissions for each of the 23 consonantal phonemes was computed and then summed across all 23 phonemes. For example, if /s/ was produced as [t] and [θ], these substitutions would be considered two different substitutions and the quantity 2 would be added to the number of different phoneme substitutions/omissions for each of the other 22 consonantal phonemes. The authors noted the consistency index does not indicate accuracy of phoneme production, only consistency of production; that is, children could have a low error consistency index because they consistently used the same, incorrect phoneme substitution across all productions of a given phoneme.

One of the most significant differences between these measures of error consistency is the type of productions compared when determining consistency. In studies of disordered child speech, the most common use of the term 'inconsistency' refers to changes in production of a single phoneme across words. Thus, productions are characterized as inconsistent if the phoneme /k/ is produced as [t] in the word *cap*, but as a glottal stop in *duck*. In this case, the different productions of /k/ could be lexically based (i.e. attributed to the fact that the phoneme occurs in different words), or positionally based (i.e. attributed to the fact that production of initial /k/ is systematically different than final /k/). If /k/ is produced as [k] in *call* but as [t] in *key*, the difference could be lexically based (as above) or could be attributed to the influence of the following vowel. Confounds such as differing lexical items and phoneme positions are problematic for analysing error consistency across multiple productions of different words because they introduce potential sources of variation other than articulatory inconsistency.

Schumacher, McNeil, Vetter, and Yoder (as cited in Hall *et al.*, 1993), Dodd (1995), Shriberg *et al.* (1997), and Ingram (2002), examined error consistency across multiple productions of the same word, rather than a single phoneme, thereby eliminating concerns about differential effects of phonetic context. Although this method of measurement controls for word structure, the various productions did not always occur in the same phrase. For example, Shriberg and colleagues and Ingram used spontaneous speech samples to obtain multiple productions of a word. Consequently, other processes such as word retrieval and grammatical encoding may have differed each time the word was produced. A precise measure of error consistency should calculate consistency based on multiple productions of the same word and in the same phrasal context.

Another limitation of the previous formulas is the denominator used as the basis for calculating percentages of error consistency. The definitions of error consistency proposed by Dodd (1995), Shriberg *et al.* (1997) and Ingram (2002) use the total number of productions as a denominator; thus, consistency is measured as a percentage of the total occurrences of a word. However, since these measures are intended to indicate *error* consistency, not *production* consistency, another approach would be to use the number of erred productions, instead of the number of overall productions, as the denominator for computing consistency.

A final difference among the previous studies is whether consistency was computed as a function of the frequency of different errors produced, as in Schumacher, McNeil, Vetter and Yoder (as cited in Hall *et al.*, 1993); Dodd, 1995; Ingram, 2002; Tyler *et al.*, 2003, or the frequency of specific error types, as in Shriberg *et al.*, 1997. Both of these types of computation provide valuable information regarding consistency. Basing error consistency on the number of different errors provides a measure of overall variability across types of errors. Computing consistency based on the frequency of specific error types, such as the most common error type (e.g. Shriberg *et al.*, 1997), reveals how frequently a child uses a given error type. A more complete analysis of error consistency would include both types of measures.

The goal of the present study is to develop a set of procedures for measuring error consistency. If consistency is to be considered a diagnostic characteristic of speech disorders, there must be objective, valid measures of the degree to which a child's productions are consistent. The formulas presented in this study improve on previous investigations of error consistency by computing consistency based on multiple productions of the same word in the same phrasal context, by calculating consistency as a percentage of the erred productions made by a child, and by including measures of both overall error frequency and frequency of the most common error type.

Methods

Participants

Data from three participants in a larger study (Betz, 2000) are presented here to illustrate the use of the three measures of consistency. The children passed a hearing screening, lived in monolingual, English speaking homes, had normal receptive language abilities as measured by a score no greater than one and a half standard deviations below the mean on the *Test of Auditory Comprehension of Language-Revised* (*TACL-R*) (Carrow-Woolfolk, 1985) and the *Peabody Picture Vocabulary Test-III* (*PPVT-III*) (Dunn and Dunn, 1997), and performed within normal limits on the muscle strength and range-of-motion portions of a structural-functional exam (e.g. no dysarthria or craniofacial anomalies).

The three participants used to illustrate the consistency measures represent a range of articulatory abilities: one child, DAS 1, was diagnosed as having developmental apraxia of speech; one child, PD 1, had a functional phonological delay; and one child, TD 1, was typically developing in terms of speech development. A certified speech-language pathologist subjectively classified the DAS and PD participants according to the criteria suggested by Davis *et al.* (1998). The DAS and PD participants scored below the 16th percentile on the *Goldman-Fristoe Test of Articulation* (*GFTA*) (Goldman and Fristoe, 1986) while the TD participant scored above the 16th percentile. Table I lists the intake characteristics of the three participants whose data are used to illustrate the consistency measures.

Table I. Participant intake characteristics

	Age	Gender	*PPVT-III*[a]	*TACL-R*[b]	*CELF-P* (recalling sentences subtest)[c]	*GFTA* (percentile score)	*GFTA* (# of errors)[d]	Speech Intelligibility (1–7 Scale)[e]
DAS 1	4;2	F	92	−0.33	5	1	40	6
PD 1	5;10	M	115	−0.31	6	8	23	3.5
TD 1	5;0	M	112	0.18	11	81	2	1

[a]Standard score, M=100, SD=15; [b]Standard score, M=0.00, SD=1.0; [c]Standard score, M=10, SD=3 (Wiig, Secord and Semel, 1992); [d]Total possible errors=73; [e]Speech intelligibility scale: 1=No noticeable differences from normal; 7=Unintelligible.

Error consistency task

An elicited production task was used to obtain multiple productions of five target words in three utterances of different lengths. The five target words, *zebra*, *spoon*, *watch*, *squirrel*, and *kite*, were chosen to include a variety of phonemes, syllable types, and word structures. Two practise stimuli, *car* and *flag*, were used to demonstrate appropriate responses. The words were elicited first in isolation and then after two carrier phrases varying in length: 'It's a _____' and 'It's a very big _____'.

Isolated words were elicited by asking the child to name picture cards. The task was introduced using the two practise stimuli. Once the child consistently named the practise stimuli in isolation, the target words were introduced. Each target was elicited four times in random order by showing the child a picture of the target word. If the child did not know the name of a picture or forgot the picture's name, the experimenter reminded the child of the target word and asked the child to repeat the word. No imitated responses were used in the analyses.

Next, target words were elicited after the short carrier phrase, 'It's a _____'. The experimenter modelled the appropriate response using the practise targets. Then, the participant repeated the experimenter's use of the carrier phrase and teaching target. Finally, the child was asked to spontaneously use the carrier phrase and the given practise word. Once the child consistently used the carrier phrase with the practise words, the target words were introduced. Each target word was elicited four times in random order. The same picture cards that were used to elicit isolated word productions were used in this condition.

Finally, target words were elicited after the long carrier phrase, 'It's a very big _____'. Children were shown enlarged picture cards of the same target and practise words. The same procedure that was used to teach the short carrier phrase was used for this condition. The procedure for eliciting target words in isolation, a short carrier phrase, and a long carrier phrase was conducted twice for each child. This procedure yielded a total of eight productions of each word in each utterance condition per child.

Data analysis

Six of the eight productions of each target word in each utterance condition were selected for analysis. Target words were rejected for analysis if there was extraneous noise during the production (e.g. if the child bumped the microphone, hit the table, or the experimenter was talking) or if the child was laughing, coughing, or hiccupping while speaking. If more than six productions were suitable, the first and/or last productions were removed from analysis.

Target words were digitized from videotape to the *Computerized Speech Laboratory 4300B* (Kay Elemetrics Corporation, 1994) at a sampling rate of 20,000 Hz. During digitization, carrier phrases were edited leaving only the target word. Then the tokens were transferred to *Multi-Speech* software (Kay Elemetrics Corporation, 1996). Two graduate students in speech-language pathology individually transcribed all target words using broad phonetic transcription. The presentation of tokens was randomized across participants, words, and utterance condition.

Reliability

Interjudge and intrajudge reliability measures were obtained on 10% of the data. Intrajudge transcription reliability was obtained using point-to-point agreement. Mean transcription reliability for judge 1 was 91% and for judge 2 was 88%. Interjudge reliability was 85%. Shriberg and Lof (1991) found that typical intra and interjudge reliability for broad phonetic transcription ranged from the mid-60s to mid-high-90s. Thus, the intrajudge and interjudge reliability for both judges fell within the accepted range for phonetic transcription.

Error consistency formulas

Error consistency was evaluated using three formulas, each measuring a different aspect of consistency. Table II describes each of the three formulas. In computing these formulas, the results were first obtained for each of the five target words in each utterance length condition and then averaged across all target words in each condition. Calculation of the number of errors made was based on the word level, not the phoneme level. That is, each formula was calculated by comparing the whole word productions of each target. If the child's production contained at least one phoneme in error it was considered an erred production. The number of individual phonemes erred in the same word was not calculated. In determining consistency in Formulas 2 and 3, two erred productions were considered 'different' if the pronunciations differed by one or more phonemes.

Formula 1 measures the proportion of total productions that were erred. Although Formula 1 does not directly assess error consistency, it provides a general impression of the overall accuracy of the child's articulation of a particular word. Formula 1 is important to calculate because the number of erred productions serves as the reference (i.e. the denominator) for Formulas 2 and 3. The possible values of Formula 1 range from 0 to 100%. A score of 0% indicated that the child made no errors on the target word and a score of 100% was given to a child who incorrectly produced all six attempts at the target. The more errors a child produced, the greater the value of Formula 1.

Table II. Error consistency formulas

Formula 1: Proportion of errors	(# errors/# total productions) × 100
Formula 2: Overall consistency of error types	(1 − (# different error types/# erred productions)) × 100
Formula 3: Consistency of the most frequently used error type	((# productions of the most frequently used error type − 1)/(# erred productions − 1)) × 100

Formula 2 measures the overall consistency of error types by comparing the number of different error types in the production of a target word to the overall number of erred productions of that word. Table II shows the procedure for calculating Formula 2. When determining the number of distinct error types, if the child used only one error type (as in TD 1's productions of *squirrel*) it was not considered a 'distinct error type' because there was no other error type to compare it to. That is, if a child used only one error type, the consistency of that error type is 100%. However, when a child uses two error types, these error types can be distinguished from one another, therefore, these errors are considered to be two distinct error types.

The number of different errors was divided by the number of erred productions, and the resulting value was subtracted from 1 to ensure the results measured consistency (i.e. higher scores reflected the use of fewer error types). The entire quantity was then multiplied by 100 to obtain a percentage score. Values for Formula 2 ranged from 0 to 100% with higher scores indicating a greater degree of error consistency. A score of 0% indicates that every erred production of a given word represents a different error type. A score of 100% corresponds to the use of only one error type for the target word. If no errors were made on any of the six productions of a word (i.e. the result of Formula 1 was 0%), then Formula 2 was not computed for that target word because it would be meaningless to discuss the consistency of error types when no errors were made.

Formula 2 is similar to the calculations used by Ingram (2002) and Dodd (1995) to quantify variability of productions. Ingram computed whole-word variation as the number of distinct forms used divided by the number of productions. Dodd's 25-word inconsistency test measured whole-word inconsistency as different productions of the same word in two out of three trials. Formula 2 differs from these previous measures because it compares the number of different error types to the number of erred productions not the number of total productions. Therefore, Formula 2 captures the literal meaning of the phrase 'error consistency' because the consistency of errors is compared to the total number of errors.

Formula 3 measures the consistency of the most frequently used error type by comparing the frequency of the most frequently used error type to the overall number of erred productions. To ensure the results of Formula 3 ranged from 0 to 100%, the quantity 1 was subtracted from both the numerator and denominator. Higher scores indicated a more consistent use of the most frequently used error type. The maximum value for Formula 3, 100%, indicated that the child used the same error type for all erred productions. A score of 0% indicated that the most frequently used error type was used only one time; that is, there was no 'most frequent' error type. As with Formula 2, values were not calculated for productions of children who did not make any errors on a specific target word because the intent of Formula 3 was to quantify the consistency of error types.

Formula 3 is motivated by Shriberg *et al.*'s (1997) error consistency calculation which measured the percentage of tokens of a given word that contained an occurrence of the most frequent error type for that word. Unlike Shriberg *et al.*'s computation, Formula 3 compares the frequency of use of error types to the number of erred productions, not the total number of productions.

Application of the error consistency formulas

Results are discussed for the three participants for each formula. Table III gives the phonetic transcription for the six productions of each target word in the isolated word

Table III. Target word productions

Participant	Target Word	Production 1	Production 2	Production 3	Production 4	Production 5	Production 6
DAS 1	kite	kætə	gæi	kæt	kætə	kæt	kætə
	spoon	əfunə	fun	fun	fuən	funə	fun
	squirrel	goza	got	gʊdə	kəlʌlə	kəʒ	kuədə
	watch	vwat	wat	wat	waʃə	vwaæ	waʃ
	zebra	tʃibə	ʒibə	ʤibæ	tibæ	ʤizba	ʤibæ
PD 1	kite	kaɪt	kaɪt	kaɪt	kaɪt	kaɪt	kaɪt
	spoon	pun	pun	pun	pun	pun	pun
	squirrel	kwowə	kwowə	kwowə	kwowə	kwowə	kwowə
	watch	watʃ	watʃ	watʃ	watʃ	watʃ	watʃ
	zebra	dibwʃ	givwə	tivwə	ʤibwə	ʤivwə	dzibwa
TD 1	kite	kaɪt	kaɪt	kaɪt	kaɪt	kaɪt	kaɪt
	spoon	spun	spun	spun	spun	spun	spun
	squirrel	skɚl	skɚl	skɚl	skɚl	skwɚl	skɚl
	watch	watʃ	watʃ	watʃ	watʃ	watʃ	watʃ
	zebra	zibrə	zibrə	zibrə	zibrə	zibrə	zibrə

condition. The results for the three formulas based on these transcriptions are given in table IV. In order to illustrate how the three formulas are computed, examples of how to calculate each formula are given in the text.

Formula 1: proportion of errors. The first calculation was Formula 1: the proportion of errors. Because the total number of productions of each target word in each utterance condition was always six, results of Formula 1 were only affected by the number of errors a child produced. A large number of errors increased the value of Formula 1. For example, DAS 1 incorrectly produced all six productions of *kite* and received a score of 100%. PD 1 correctly produced all six utterances of *kite* corresponding to a score of 0%. TD 1 had five erred productions of the target *squirrel* resulting in a score of 83% for Formula 1.

Table IV. Consistency results

Participant	Target Word	Formula 1	Formula 2	Formula 3
DAS 1	kite	100	50	40
	spoon	100	33	40
	squirrel	100	0	0
	watch	100	17	20
	zebra	100	0	0
PD 1	kite	0	n/a	n/a
	spoon	100	100	100
	squirrel	100	100	100
	watch	0	n/a	n/a
	zebra	100	0	0
TD 1	kite	0	n/a	n/a
	spoon	0	n/a	n/a
	squirrel	83	100	100
	watch	0	n/a	n/a
	zebra	0	n/a	n/a

n/a: Not applicable due to no erred productions.

Formula 1, DAS 1, kite: (6 errors/6 total productions) × 100=6/6 × 100=100%

Formula 1, PD 1, kite: (0 errors/6 total productions) × 100=0/6 × 100=0%

Formula 1, TD 1, squirrel: (5 errors/6 total productions) × 100=5/6 × 100=83%

Because the values for Formula 1 are percentages, results can be compared between different targets for the same child as well as between children. For example, the results for Formula 1 given in table IV show that the DAS and PD participants produced a greater percentage of errors than the TD participant.

Formula 2: consistency of error types. The second calculation, Formula 2, measured the consistency of error types produced in multiple productions of the same word. That is, the value of Formula 2 was directly related to the number of different errors a child made for a target word. Comparison of tables III and IV shows that the fewer error types a child made, the higher the child's value for Formula 2. For example, TD 1 made only one type of error on the word *squirrel*, [skɚl] (i.e. the number of different errors was 0 because there were no other errors to compare to [skɚl]), so his value for Formula 2 was 100%. For the target *kite*, DAS 1 produced three different error types, [kætə], [gæi], and [kæt], resulting in a value of 50% for Formula 2. PD 1 incorrectly produced every attempt at the word *zebra* and each production was a different error type. This pattern of errors resulted in a value of 0% for Formula 2.

Formula 2, TD 1, squirrel: (1−(0 different error types/5 erred productions)) × 100=(1−(0/5)) × 100=100%

Formula 2, DAS 1, kite: (1−(3 different error types/6 erred productions)) × 100=(1−(3/6)) × 100=50%

Formula 2, PD 1, zebra: (1−(6 different error types/6 erred productions)) × 100=(1−(6/6)) × 100=0%

These results show that higher values of Formula 2 correspond to more consistent use of specific error types and lower values indicate inconsistent use of error types. As with Formula 1, results of Formula 2 can be compared within and between participants. Examination of table IV reveals that in general, TD1 and PD1 provided more consistent error types than DAS1, however, PD1 was not consistent in his productions of *zebra*.

Formula 3: consistency of the most frequently used error type. The final measure of error consistency was Formula 3, which measured the percentage of erred productions that were the most frequent error type. To illustrate the calculation of Formula 3, first consider the errors made by TD 1 for the target *squirrel*. TD 1 incorrectly produced five of the six instances of *squirrel*, but used only one error type, [skɚl]. Therefore, [skɚl] was the most frequent error type, resulting in a score of 100% for Formula 3. DAS 1 incorrectly produced all six attempts at *kite*, and used three different error types. The most frequent error type, [kætə], was used 3 times.

Formula 3, TD 1, squirrel: ((# productions of the most frequent error type−1)/(# erred productions−1)) × 100=((5 productions of [skɚl]−1)/(5 erred productions−1)) × 100=(5−1)/(5−1)) × 100=100%

Formula 3, DAS 1, kite: ((# productions of the most frequent error type−1)/(# erred productions−1)) × 100=((3 productions of [kætə]−1)/(6 erred productions−1)) × 100=((3−1)/(6−1)) × 100=40%

Comparison of tables III and IV shows that values for Formula 3 are higher when the child is more consistent at using one error type.

Discussion

Consistency measures

The methods presented in this study provide a framework for researchers investigating error consistency. Few researchers in this field have operationally defined 'error consistency' or stated the context or formula used to measure consistency. In order to account for different types of error consistency, three formulas were developed to quantify various aspects of this phenomenon. The formulas measured the proportion of errors, overall consistency of error types, and consistency of the most frequently used error type across multiple repetitions of the same target word.

It is important to note that, in the larger study from which the examples were drawn, the potential effects of phrase length and linguistic context on error consistency were eliminated by eliciting target words in three utterance conditions: in isolation, after a short carrier phrase, and after a long carrier phrase. This approach improved upon previous studies that computed consistency by comparing phonemes from different words occurring in different contexts. Because target words were elicited in a structured manner, error consistency was measured across multiple productions of the same word in a controlled set of syntactic contexts.

The first formula presented, Formula 1, measured the proportion of errors made by a child. The score a child received for this formula indicated the percentage of productions that were erred. According to Schumacher, McNeil, Vetter and Yoder (as cited in Hall *et al.*, 1993), the number of erred productions of a given word is 'consistency'. However, in this sense 'consistency' only indicates whether a child incorrectly produces the same word over multiple trials, not whether the child makes different errors each time it is produced. Therefore, this formula alone should not be used as a measure of consistency. However, Formula 1 is useful because it provides a general impression of the child's overall accuracy.

Formula 2 measured the consistency of error types produced by a child. A child's score reflected how many different errors the child made across repeated productions of the same target. The presence of different errors on repetitions of the same word has also been referred to as 'variability of errors' (Schumacher *et al.*, as cited in Hall *et al.*, 1993). This definition of variability is probably what researchers and clinicians intend when they state that a child makes 'inconsistent errors'. However, consistency of error types alone does not account for all of the possible differences between two children. As an example, suppose two children each produce a target word six times and each uses two different pronunciations for the target. Child one uses each pronunciation three times and child two uses one of the pronunciations once and the other five times. Both of these children would receive a score of 67% for Formula 2. However, it would be misleading to say that each child has the same degree of consistency.

Formula 3 measured the consistency of the most frequently used error type. Scores on this formula revealed how often the child used the most frequent erred production for a given target word. Take the example of the two children mentioned above. Each would have received the same score for Formula 2, but for Formula 3, child one would score 40% and child two would score 80%. One might think that since Formula 3 differentiated the two children and Formula 2 did not, Formula 3 is a better measure of consistency. This conclusion is not necessarily true. Consider two other children who each say a target word six times and both use their most commonly used production three times. One child uses the same pronunciation for the remaining three productions of the word, but the other child makes three different errors. Both of these children would receive the same score, 33%, for Formulas 3, but the first child would receive a higher score on Formula 2 because he used a greater number of different error types. Thus, Formulas 2 and 3 are both useful in determining error consistency.

Consistency in this study was measured at the word level; however, the formulas can easily be used to compute consistency at the phoneme level as well, by counting the number of different productions of a given phoneme rather than the number of different productions of the word. However, as in this study, these calculations should only be computed across multiple productions of the same word to rule out effects of the target words and of word position on consistency measures.

Differential diagnosis of DAS and PD

These three formulas were developed as part of a study (Betz, 2000) investigating error consistency in children with DAS and PD. Eleven children participated in this study, five children with DAS, five children with PD, and one typically developing child. The participants met the same subject selection criteria as the three children discussed in detail earlier and all the participants completed the experimental task as described in the methods section. It was hypothesized that children with PD would produce more consistent errors than children with DAS. This hypothesis was based on previous research citing 'inconsistency of errors' as a characteristic of children with DAS. However, few researchers included a control group of children with PD or mathematically calculated the degree of consistency in children with DAS. Based on a series of Mann-Whitney *U* tests, the study found significant differences between the children with DAS and the children with PD for Formula 1 but no statistical difference for Formulas 2 or 3. These results indicated that although children with DAS produced more errors than the children with PD, the consistency of errors did not differ. Because of the small sample size in this study, these results should be interpreted with caution; one reason for the lack of significant results for Formulas 2 and 3 may be the lack of power to identify real differences.

Another possibility for the lack of significant results for Formulas 2 and 3 may be that error consistency is related not to the presence of DAS *per se*, but to severity of the disorder. The original study investigated the relationship between error consistency and severity of speech disorder. No statistically significant results were found for Formulas 2 or 3; however, the four children with the most severe speech disorder (i.e. lowest intelligibility ratings and highest number of errors on the *GFTA*) produced the least consistent errors. Three of these four children were diagnosed as having DAS and one as having PD. These findings suggest that clinicians may be diagnosing DAS in all children with severe speech disorders.

The original study also examined the effects of phrase length on error consistency. It was hypothesized that children with DAS might have decreased error consistency in the longer utterance length conditions because these children are known to have increased difficulty with longer utterances. The results showed that error consistency, as measured by Formulas 1, 2 and 3, did not differ across the three utterance length conditions (i.e. isolated words, short carrier phrase and long carrier phrase). The lack of significant results may have been due to the repeated use of the same carrier phrases in which the target word always occurred in phrase final position. These phrases may have become routine, decreasing the need for linguistic and motor planning. Additionally, because the target word was in phrase final position, the children did not have to continue planning the next part of the utterance while simultaneously producing the target. If the target words were embedded in carrier phrases, additional linguistic and motor planning would need to occur during and after the target word production potentially affecting error consistency.

The three formulas described in this study can be used by researchers interested in error consistency. They may be useful in investigations of consistency as a predictor of change over time and in studies of the characteristics of various types of phonological disorder. To date, these formulas have only been used in a small-scale study of eleven participants. Future research is needed with children with DAS and PD to determine whether error consistency is a defining characteristic of DAS.

Acknowledgments

This research was conducted as part of the first author's master's thesis at the University of Washington. Preliminary report of this study was made at the 2000 ASHA convention in Washington, DC. We thank Edy Strand for help with planning this study, Amy Skinder Meredith for help with data collection and Jill Beilinson and Ann Sweet for their help with transcription. We also thank the Seattle area schools and speech-language pathologists for their help with recruiting participants and the children and their families for their participation.

References

Betz, S. K. (2000). Articulatory Error Inconsistency in Children with Developmental Apraxia of Speech and Functional Phonologic Delay. Unpublished master's thesis, University of Washington, Seattle, WA.

Bradford, A., & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics and Phonetics*, *10*, 77–101.

Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language-Revised*. Allen, TX: DLM.

Davis, B., Jakielski, K., & Marquardt, T. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics and Phonetics*, *12*, 25–45.

Davis, B., & Velleman, S. (2000). Differential diagnosis and treatment of developmental apraxia of speech in infants and toddlers. *Infant and Toddler Intervention*, *10*, 177–192.

Dodd, B. (1995). Procedures for classification of subgroups of speech disorder. In B. Dodd (Ed.), *The Differential Diagnosis and Treatment of Children with Speech Disorder* (pp. 49–64). San Diego, CA: Singular Publishing Group.

Dodd, B., & McCormack, P. (1995). A model of speech processing of phonological disorders. In B. Dodd (Ed.), *The Differential Diagnosis and Treatment of Children with Speech Disorder* (pp. 65–89). San Diego, CA: Singular Publishing Group.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test-III*. Circle Pines, MN: American Guidance Service.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439.

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology*, *12*, 376–380.

Forrest, K., Dinnsen, D., & Elbert, M. (1997). Impact of substitution patterns on phonological learning by misarticulating children. *Clinical Linguistics and Phonetics*, *11*, 63–76.

Goldman, R., & Fristoe, M. (1986). *Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.

Hall, P., Jordan, L., & Robin, D. (1993). *Developmental Apraxia of Speech: Theory and Clinical Practice*. Austin, TX: Pro-Ed.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language*, *29*, 713–733.

Kay Elemetrics Corporation. (1994). *Computerized Speech Lab Model 4300B*. New Jersey: Kay Elemetrics.

Kay Elemetrics Corporation. (1996). *Multi-Speech Signal Analysis Workstation Model 3700*. New Jersey: Kay Elemetrics.

Shriberg, L., Aram, D., & Kwiatkowski, J. (1997). Developmental apraxia of speech II: Toward a diagnostic marker. *Journal of Speech, Language and Hearing Research*, *40*, 286–312.

Shriberg, L., & Lof, G. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, *5*, 225–279.

Sosa, A. V., & Stoel-Gammon, C. (2003). Intra-word phonological variability during early lexical acquisition. Child Phonology Conference 2003, University of British Columbia, Vancouver, BC, 1–4 July.

Tyler, A., Lewis, K., & Welch, C. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology*, *12*, 289–298.

Wiig, E., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals-Preschool*. San Antonio, TX: The Psychological Corporation.