
Letters to the Editor

Comments Regarding the Investigation of Developmental Apraxia of Speech: Response to Shriberg, Aram, and Kwiatkowski

Several concerns prompt me to write regarding the series of articles concerning Developmental Apraxia of Speech (DAS) by Shriberg, Aram, and Kwiatkowski (1997a, 1997b, 1997c). Although I question the relevance of the authors' goal of identifying either a *single* or *unitary* diagnostic marker for DAS based on speech error profiles or a potential phenotype marker, I am more troubled by several methodological issues.

Method

Subjects

The subject selection in the series is flawed. The only criterion for the selection of the first sample of 14 subjects was that they had "previously been seen in a study series on developmental apraxia of speech" (Shriberg et al., 1997b, p. 287). The authors stated that the Speech Disorders Classification System (SDCS) provides qualitative support for why these children are suspected to have DAS" (p. 290). I strongly disagree. If I interpret the SDCS classification entries in Table 2 (Shriberg et al., 1997b) correctly, then the sample of children with DAS contained: (a) three children with speech delay, (b) five children who had speech delay plus an error pattern that includes uncommon clinical distortions in over 20% of the words in a sample, (c) one child who had normal or normalized speech with errors acceptable for age, (d) two children who had both common clinical distortion errors and continuingly imprecise speech, (e) one child who had marginally imprecise speech alone, and (f) one child who had normal or normalized speech with errors acceptable for age, plus an error pattern that includes uncommon clinical distortions in over 20% of the words in a sample.

None of these speech disorder classifications provide enough information to consider a diagnosis of DAS. The authors provided no information about nonsegmental aspects of the subjects' speech, and no information about the subjects' response to treatment. Even if the authors had been able to summarize the nature of the subjects' speech disorder clearly, they completely failed to describe why and how the children with suspected DAS were selected for the original study series. The descriptive data about the first sample of 14 subjects show that these subjects vary widely in terms of severity levels, age, and intelligence quotients (see Tables 1 and 2, Shriberg et al., 1997b).

The only criteria for selecting the children without DAS in the comparison group were that they had developmental phonological disorders of unknown origin (for the 64 younger children) and a speech disorder persisting past the age of 7 years and no indication that the speech delay might be associated with DAS (for the older group).

The second study, referred to as an internal cross-validation study, included an additional 20 children, 14 of whom who were referred to the University of Wisconsin Phonology Clinic because of a referring clinician's "concern not only with slow progress, but also with a specific observation that raised the question of possible or suspected DAS" (Shriberg et al., 1997a, p. 281). Who made the observation that raised the question of possible DAS, the referring clinician, or one of the authors? What specific observations raised the question? The other 6 children in the second study were simply selected because they were referred to the Phonology Clinic and had "suspected DAS."

The authors' failure to clearly describe the criteria for membership in the research and comparison groups undermines the validity of the subsequent descriptions, analyses, and conclusions. Soliciting speech samples of children with suspected DAS from other researchers was a far more objective and creative means of identifying subjects (Shriberg et al., 1997c), but the researchers still went to the trouble of comparing this group of subjects to an ill-defined comparison group.

Procedures

In my opinion, the authors' attempt to quantify the consistency of error failed to capture the *inconsistency* and *variability* in speech production that has been associated with DAS (Hall, Jordan, & Robin, 1993; Murdoch, Porter, Younger, & Ozanne, 1984; Rosenbek & Wertz, 1972). Although the authors provide an example of computing the "per child consistency percentage" (p. 299), it is difficult to interpret the measurement. What does a "per child error consistency percentage" of 50% mean? Does it mean that half of the child's errors are consistent? Does it mean that half the time that the child attempts to say *dog*, the child instead says *dod*? Although Shriberg et al. (1997b) might have addressed the consistency of their subjects' most frequently occurring errors in a conversational speech sample, they did not address the variability of the errors in the sample. In my experience, the inconsistency associated with DAS is a treatment issue, not a cardinal symptom of spontaneous conversational speech. The more remarkable inconsistency issue is the variability of children's attempts to volitionally repeat an utterance several times in the structure of a treatment session (Hall et al., 1993).

Results and Interpretation

Although the tables in all three articles in this series displayed data clearly and concisely, the figures are difficult to interpret. For example, Figure 5 (Shriberg et al., 1997b, p. 300) and Figure 2 (Shriberg et al., 1997c, p. 321) appear to display similar data regarding error consistency for two different groups of children with DAS. However, Figure 5 displays consonant error consistency for all the children in Study I, where Figure 2 displays data only for the children with DAS in Study III who were judged to have inappropriate stress.

The authors made several summary statements that contradicted or ignored the data presented in their own tables and figures. For example, Table 4 (Shriberg et al., 1997b, p. 293) and Table 3 (Shriberg et al., 1997c, p. 317) display Speech Severity Measures which show that the younger children with suspected DAS have significantly lower Intelligibility Indices than the age-matched comparison children with speech delay. Table 4 (Shriberg et al., 1997b) also shows significant differences in Percentage of Consonants Correct and Percentage of Vowels Correct for the younger children with suspected DAS. The authors concluded, "Added to the findings that none of the speech severity measures replicated significant findings for both younger and older comparisons, these analyses suggested that severity of involvement was not a viable candidate for a diagnostic marker of DAS" (Shriberg et al., 1997b, p. 295). This ignores the significance of the differences in intelligibility, the most global of these speech severity measures, and a likely indication of the communicative impact of the child's speech disorder. These measures could be interpreted to mean that *intelligibility*, not severity, might be a discriminating diagnostic feature for some younger children, who may not have had any or as much treatment as older children.

The authors' conclusions about speech rate are also confusing. Prosody-voice data for Study I children are presented in Figure 1 (Shriberg et al., 1997b). The data show that rate, stress, and resonance were rated as appropriate in a significantly lower percentage of utterances for younger children with suspected DAS compared to children with speech delay (Shriberg et al., 1997b). For Study II children, data concerning prosody-voice analyses are presented in Table 7 (Shriberg et al., 1997b), but in different form. This table shows that 4 of 12 younger children and 5 of 8 older children (9 of 20 combined) had rates judged as inappropriate or questionable. In the third study, prosody-voice data were presented in Figure 1 (Shriberg et al., 1997c). The younger children with suspected DAS were again shown to have significantly lower average scores on both rate and stress. As with the Speech Severity Measures, the authors chose to ignore the consistently different findings

regarding rate for the younger children. Why were these striking differences in the younger children not addressed?

Summary

The conclusions and discussion were based on only 16 children who were identified based on clearly described criteria. Those subjects have widely varying ages, unspecified treatment histories, and varying receptive language status. The authors did not claim that these results are generalizable to a larger population of children, but I fear that their results will be interpreted that way by others. Identifying unusual phrasal stress deficits in 8 of 16 children with suspected DAS should not be construed or implied as evidence for a diagnostic marker for a subtype of DAS.

I strongly agree with the authors that longitudinal studies of children with persistent and unusual speech disorders are needed. Studies concerning children's responses to treatment are also needed. Although models of adult onset apraxia (AOS) may provide useful procedures for measuring or describing speech and nonspeech characteristics of DAS, using AOS as a theoretical model or clinical analogy to DAS leads us to ask less relevant questions about children with unusual and persistent speech disorders. Children with suspected DAS are different from adults who have AOS. Children who have never spoken normally or used language normally are different from adults who have acquired a speech disorder after decades of using spoken and written language normally. In order to intervene efficiently and appropriately, we need to know whether and how children with DAS differ from other children, not how they might resemble adults with an acquired disorder.

Claire M. Waldron
Radford University
Radford, Virginia

References

- Hall, P. K., Jordan, L. S., & Robin, D. A. (1993). *Developmental apraxia of speech: Theory and clinical practice*. Austin, TX: Pro-Ed.
- Murdoch, B. E., Porter, S., Younger, R., & Ozanne, A. (1984). Behaviors identified by South Australian clinicians as differentially diagnostic of developmental dyspraxia. *Australian Journal of Human Communication Disorders*, 12, 55-70.
- Rosenbek, J. C., & Wertz, R. T. (1972). A review of fifty cases of developmental apraxia of speech. *Language, Speech, and Hearing Services in Schools*, 5, 23-33.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997a). Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research*, 40, 273-285.

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

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997c). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, 40, 313-337.

Received June 30, 1997

Accepted February 16, 1998

Alternative Research Perspectives: A Response to Waldron

We appreciate Waldron's interest in our work and welcome the opportunity to clarify goals, methods, findings, and interpretations.

Goals

Waldron begins by questioning "the relevance of the authors' goal of identifying either a single or unitary diagnostic marker for DAS (Developmental Apraxia of Speech) based on speech error profiles or a potential phenotype marker." There certainly are alternatives to the medical model underlying this work. Unfortunately, because Waldron elects not to specify her concern, we cannot take this opportunity to address her apparently strong reservations about the relevance of our research questions.

Methods

Participants

Waldron asserts that "the subject selection in the series is flawed." Later in these five paragraphs she concludes that "the authors' failure to clearly describe the criteria for membership in the research and comparison groups undermines the validity of the subsequent descriptions, analyses, and conclusions." Technically, she appears to claim that our data lack internal validity, external validity, or both, due to lack of appropriate inclusionary and exclusionary criteria for the DAS, non-DAS, or both groups, and some unspecified heterogeneity problem with associated subject characteristics.

We have difficulty interpreting the source of Waldron's concerns: Does she question the adequacy and clarity of our subject descriptions, or the rationale for our inclusionary and exclusionary criteria? As we provide considerable information in tabular form on all

subjects, and additional information is available in publications we cite, we suspect her reservations primarily reflect the latter—a fundamental disagreement with our research perspective and consequent research design. Briefly, the perspective and its implementation in our research design are as follows.

In the first paper of the series (Shriberg, Aram, & Kwiatkowski, 1997a), findings from a review of the literature and a local ascertainment study led us to conclude that *suspected DAS* is the only appropriate term for this putative clinical entity. Accordingly, we suggested that the only defensible inclusionary criterion for suspected DAS is referral by a competent clinician who has worked with a child who (a) has failed to make progress (i.e., normalize) and (b) has some characteristics found on typical DAS checklists. For these diagnostic marker studies we deliberately did not set arbitrary inclusionary criteria taken from the many available DAS checklists. Rather, given the diversity of perspectives in the literature, we proposed that DAS is not currently defined by the results of any particular battery of case history and test findings, nor does diagnosis by any one DAS researcher gain any additional validity over the judgments of an experienced clinician. We discussed the problem with the circularity of this reasoning in the design of DAS research, as have many prior reviewers of this literature.

Thus, our inclusionary criterion for the three studies was a referral for suspected DAS by an experienced Madison-area clinician (local studies), an experienced DAS clinical researcher (second author; validation study), and five additional experienced DAS clinical researchers (cross-validation study), each of whom used his or her own inclusionary criteria for suspected DAS. These criteria were thus purposely unconstrained in order to include in these studies children with the range of features for suspected DAS described in the literature. For optimal external validity, as well, our exclusionary criterion for children with speech delay was expressly limited only to children with no indication of suspected DAS. We provided citations to other readily accessible papers that include additional details on our methods and definitions for children with speech delay of unknown origin.

From this perspective, we have difficulty with Waldron's perception of an "ill defined" control group. Waldron might reasonably presume that we fail to find a common speech characteristic for children with suspected DAS (i.e., a potential Type II error) because we did not include, exclude, or both, the appropriate subjects in each group, based on both speech and socio-demographic variables. But note that we *did* identify a possible diagnostic marker amidst this diversity—inappropriate stress. Alternatively, perhaps the concern is that these procedures created an internal threat to the

validity of the potential stress marker (i.e., a possible Type I error). From the general tone of her response, this appears to be the focus of Waldron's concern but one she unfortunately does not specifically address. For a helpful dialogue on this issue we would ask Waldron to specify exactly which inclusionary and exclusionary subject variables for both speech groups she would recommend, and specifically how the absence of these criteria invalidates our findings. Further, the comment on our use of the Speech Disorders Classification System (SDCS) findings appears to miss the point. As described in the paper, the uncommon distortion errors and imprecise speech described in SDCS codes for many of the children provide only qualitative support for the experienced clinician's use of the term suspected DAS; these descriptive data were not themselves used as inclusionary criteria.

Consistency and Variability

Waldron's observations on her clinical experiences with inconsistency and variability in a treatment context are interesting and warrant study in repeated measures designs using controlled sentence repetitions. However, the data set in this relatively large-scale collaborative study consisted solely of available recorded conversational speech samples. We provided detailed rationale for the error consistency approach with these samples, rather than error variability approach (i.e., insufficient tokens were available for the latter), including the formula for error consistency and a complete example. We are puzzled by Waldron's question about interpretation of the metric. As described in the paper, a consistency score of 50% indicates that 50% of the tokens on eligible word types in conversational speech involve similar substitutions. Statistical tests of the error consistency scores indicated "no significant between-groups differences in error consistency for consonants as assessed at the level of individual sounds and as aggregated over total sounds, three developmental sound classes, two class features, two voicing features, and five manner features" (Shriberg et al., 1997b, p. 301). Thus, findings indicated that, as assessed in conversational speech samples, error consistency scores do not significantly discriminate children with suspected DAS from children with speech delay.

Results and Interpretation

Waldron makes the unconstrained claim that "the figures are difficult to interpret" in the three articles. To support the claim, she describes her difficulty understanding two figures in different papers in which the same dependent variables are averaged for different

subgroups of children. Possibly the confusion is more in tracking our lengthy analyses than with the figures themselves. As described in the third paper, the latter figure occurs in the context of an analysis of only those children with inappropriate stress aggregated over data sets. The data for the two figures in question address different questions and do not conflict.

Waldron's second concern is that we "contradicted or ignored the data presented in [our] own tables and figures," specifying our evident disregard of the intelligibility findings. As described in our research rationale, between-groups analyses comprise only the first step in identifying a potential diagnostic marker. For each potentially discriminating factor based on the statistically significant between-groups findings, per-child data are needed to determine whether a specific performance level can be used to discriminate group membership. We reviewed all such findings under the heading "Individual Subject Analysis." For the Intelligibility Index findings in Table 3, we described this secondary analysis of the speech data as follows: "Scatter plots of scores of children in the DAS and SD (Speech Delay) groups indicated overlapping values at all levels of the total and subdivided indices. Thus, as found in Study 1, *severity of speech involvement* (a cover term in Table 3 and elsewhere for all the speech metrics, including the Intelligibility Index) was not a discriminating diagnostic feature" (Shriberg et al., 1997c, p. 317).

Waldron's third claim is that our conclusions about speech rate are confusing, intimating that we chose to ignore the rate findings for the younger children: "Why were these striking differences in the younger children not addressed?" In fact, as above, we did thoroughly address these findings. Follow-up individual analysis of all the between-groups, prosody-voice findings failed to support rate as a potential diagnostic marker. In Shriberg et al. (1997b, p. 295) we state, "In contrast to the speech findings, scatter plots for the prosody-voice findings suggested the possibility that stress and quality (i.e., not rate) might be candidates for diagnostic markers. This conclusion was based on both the scatter plot findings and on the replication across age groups." We do not agree with Waldron's characterization of the rate data as "striking" because only the first of the two statistically significant younger children comparisons yielded a theoretically and clinically significant effect size. For the second comparison, the mean difference was only approximately four percentage points, and both groups' average scores were well above the 90% criterion for appropriate rate (Shriberg et al., 1997c, Figure 1, p. 318). Because we used customary statistical significance levels without Bonferroni alpha corrections for family-wise testing, replicated findings across studies, as described in these papers, was a requirement to protect against overinterpretation of any one statistically

significant finding. Finally, on the suggestion that we ignored the rate data, the third paper in the series includes a focused discussion of the implications of the rate findings relative to alternative psycholinguistic loci of the inappropriate stress deficit observed in over half of the children with suspected DAS.

Summary

Waldron begins the final section of her letter with the claim that the data for all but 16 of the 48 children with suspected DAS should be discounted, noting that even the 16 subjects "have widely varying ages, unspecified treatment histories, and varying receptive language status." Waldron evidently views the data for the 16 children provided by the five collaborating researchers (Study 3: A–E) as valid, but not the data obtained from the samples provided by the second author (Aram & Horowitz, 1993, as cited in Shriberg et al., 1997b) or the samples provided by local clinicians (Study 2). As reviewed at the outset of the present discussion, we take a different position on both criteria for suspected DAS and on the value of subject diversity for the questions posed. What is most relevant from a scientific perspective is that even if our findings were restricted to the 16 children whom Waldron accepts as valid subjects, we submit that they would sufficiently motivate our interpretations and conclusions. The issue here appears to be a difference in perspective on the role and responsibilities of researchers. Waldron is concerned that, notwithstanding the caveats we include about threats to internal and external validity of our findings, some readers might act inappropriately in the clinic based on their interpretation of our findings. We report that inappropriate stress occurred in approximately 50% of the children with suspected DAS (whether totaling 16 or 48 children), compared to 10% of the 71 eligible children with speech delay of unknown origin. In our view, our fundamental task as researchers is to provide a clear report of our science. In turn, if sufficiently stimulating to the scientific community, the report may motivate others to attempt to replicate and explicate our findings.

Finally, we respect Waldron's interest in learning how children with DAS differ from other children, but we disagree with her conclusion that studies comparing DAS with adult onset apraxia (AOS) "leads us to ask less relevant questions about children with unusual and persistent speech disorders." Perhaps the source of our difference is in alternative perceptions of how to reach the end goal of helping children, possibly including prevention. Waldron stresses the need for treatment-relevant research, noting, "In order to intervene efficiently and appropriately, we need to know whether and how

children with DAS differ from other children, not how they might resemble adults with an acquired disorder.” We, too, are interested in treatment, but specifically as it follows from a well-developed explanatory account of the origin and nature of the disorder (see discussion, Shriberg et al., 1997c, pp. 332–333). A possible route toward such an account is to explore the relevance of neuroscience data in adult onset speech disorders, an approach that has been productive for studying dysarthria in adults and children. As reviewed in the first and third papers, our interest in how the speech of children with suspected DAS might resemble the speech of adults with acquired apraxia addresses the hypothesis of common deficits in underlying neurolinguistic or psycholinguistic processes. A report from such a study is in preparation.

Lawrence D. Shriberg
University of Wisconsin–Madison

Dorothy M. Aram
Emerson College
Boston, Massachusetts

Joan Kwiatkowski
University of Wisconsin–Madison

References

- Aram, D. M., & Horowitz, S. J.** (1983). Sequential and non-speech praxic abilities in developmental verbal apraxia. *Developmental Medicine and Child Neurology*, *25*, 197–206.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J.** (1997a). Developmental apraxia of speech: I. Descriptive perspectives. *Journal of Speech, Language, and Hearing Research*, *40*, 273–285.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J.** (1997b). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research*, *40*, 286–312.
- Shriberg, L. D., Aram, D. M., & Kwiatkowski, J.** (1997c). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research*, *40*, 313–337.

Received April 17, 1998

Accepted April 17, 1998