

the articulation test than they did in conversational speech. It is not obvious to me, however, how to interpret this trend. One could argue that these are more useful data, since substitutions show directly how children are using particular sounds within their linguistic system. But individual linguistic analyses would need to be done to see if this is the case.

In summary, Morrison and Shriberg have presented us with some important data on how children's speech may vary according to elicitation mode. They have also made an interesting case that conversational samples may be the best way to obtain data on the nature of the child's speech. They have not, however, demonstrated that articulation tests cannot be used to yield comparable results. On the contrary, their data indicate that articulation tests can provide roughly comparable data in a more efficient manner than the much more time-consuming procedure of language sampling and transcription.

David Ingram
Department of Linguistics
The University of British Columbia

References

- Ingram, D. (1989). *First language acquisition: Method, description, and explanation*. Cambridge: Cambridge University Press.
- Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, 35, 259-273.
- Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1969). *Photo Articulation Test* (2nd ed.). Danville, IL: The Interstate Publishers.
- 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.
- Templin, M. C. (1957). Certain language skills in children: Their development and interrelationships. *Institute of Child Welfare Monographs Series 26*. Minneapolis: University of Minnesota Press.

Received June 28, 1993
Accepted February 25, 1994

Response to Ingram Letter

We welcome the opportunity to address alternative perspectives on assessment in child phonology. The continuing differences on validity and efficiency perspectives raised in David Ingram's letter are the very ones that motivated our study. In the following four sections we address issues he raises and present our reasons for disagreeing with each of his perspectives.

1. Ingram's first assertion is that, although our study documents "statistical differences between the two elicitation modes on individual sounds, manner types, and phonological processes," the article does not address "implications of these differences for research and clinical purposes."

We are puzzled by this assertion because in multiple sections of the paper we address research and clinical implications of the findings. On page 267 we begin a section of the article entitled *Pass-Fail Analysis* with the following sentence: "A

series of analyses was conducted for a perspective on the types of research and clinical decisions made from findings in each of the two speech sampling modes." Among the many statistically significant and descriptive findings in the pass-fail analyses (see Table 4 and Figure 10), we demonstrate that only 17 (71%) of the consonant sounds received similar clinical decisions in the two sampling modes and that clinical outcome discrepancies between the two sampling modes for two thirds of the 61 children occurred on as few as 13 speech sounds per child to as many as 20 speech sounds per child. A series of individual difference analyses (see Table 5) document elicitation mode differences in relation to six clinical-research variables (*Age*, *Average words per utterance*, *Percent vowels correct*, *Percent consonant singletons correct*, *Percent consonant clusters correct*, and *Intelligibility*). Two discussion sections entitled *Research Considerations* and *Clinical Considerations* address a variety of additional differences between elicitation modes, including clinical and research implications of reliability issues in phonetic transcription. Clearly, the express focus of this study was to examine, in some detail, the research and clinical implications of mode of elicitation on group- and individual-level speech data using a large sample of children with developmental phonological disorders.

2. Ingram's second claim is that, notwithstanding the significant statistical differences between modes of elicitation that occurred at every level of the speech data, "there is an alternative interpretation" of the data that could be used to argue "for [emphasis added] the use of articulation tests." Ingram's one example of an alternative interpretation in support of articulation testing involves a recast of the interval-level percentage data in Figure 5 to nominal-level categorical data. Specifically, he suggests that if a researcher were interested in using articulation test results only to document which consonants children have acquired, these data would yield decisions similar to those based on conversational speech for 22 (50% criteria) and 21 (75% criteria) of the 23 consonants.

Ingram's suggested alternative interpretation does not follow from the published data. These averaged, group-level data cannot be used to draw conclusions about subject-level percentages of mastery. Customary acquisition data such as those presented in Sander's (1972) widely cited bar graph and the comprehensive study by Smit, Hand, Freilinger, Bernthal, and Bird (1990) reflect the percentages of children reaching criteria for each sound, not per-sound percentages correct data averaged over children—the form of our data. Thus, our reported data could not be used to support a claim that phonetic inventories for each of these 61 children would be similar whether based on responses to the articulation test or the continuous conversational speech sample.

Ingram also attempts to explain the possible origin of the differences in /j/ percentages obtained in each elicitation mode. He asserts that to evoke /j/ medially, the Photo Articulation Test uses the words "canyon," "union," and "lawyer." In fact, these words appear only in a list of supplementary words in the *PAT Manual*. Our data were obtained using the standard PAT procedure in which word-medial /j/ is evoked spontaneously by the more age-appropriate form, "thank you."

3. A third observation Ingram offers in support of articulation testing is based on the entries in his Table 1. What is puzzling

about this table is that the publication he cites (Shriberg, 1993) used just such data to support the construct validity of the three developmental sound classes—Early-8, Middle-8, and Late-8 sounds. Specifically, Figure 8 in Shriberg (1993) is a six-panel display illustrating the high concordance between normative data on speech-sound mastery obtained with articulation tests and the descending sort of the 23 mastery percentages from children with speech delay into the three developmental sound classes. Our argument for the construct validity of the three developmental classes as a developmental ontology was based on numerical analyses of data such as those in Ingram's Table 1, which he suggests "line up rather well." Once again, the data in Ingram's Table 1 and the detailed numerical examination of the same question available in Shriberg (1993) indicate only that there is group-level stability in averaged percentages in both elicitations. Neither examination supports Ingram's claim that the acquisition sequences obtained by each mode are, in his words, "the same."

4. A fundamental difference between our position on these clinical research data and Ingram's evident perspective concerns the scientific and ethical choices one makes when faced with two modes of data gathering, one of which has serious construct validity problems, but purportedly is more efficient. This seems to be the crux of Ingram's position, as indicated by his final two sentences: "They have not, however, demonstrated that articulation tests cannot be used to yield comparable results. On the contrary, their data indicate that articulation tests can provide roughly comparable data in a more efficient manner than the much more time-consuming procedure of language sampling and transcription." This tension between efficiency and validity is a longstanding issue in speech assessment. It is the very issue that we set out to examine closely and that we try to delineate in our review of literature, presentation of findings, and extended discussion. Here we restrict comments to the problems we have with the reasoning expressed in Ingram's letter.

Efficiency. To date, the claim that speech-language sampling takes more time than articulation testing is only an undocumented generalization based on a researcher's or clinician's individual realm of experience. Published, representative data comparing sampling times in each elicitation mode over a wide range of applications are not, to our knowledge, available. Many current word-level procedures for articulation "testing" involve lists numbering several hundred words, which when evoked spontaneously with pictures and objects, take considerable time. In our experience testing children and adults through the lifespan, conversational speech sampling often takes less time than citation form procedures, especially when testing older persons and children with minimal speech-language involvements.

For the purposes at hand, however, we will attest that for certain clinical and research questions with samples from certain populations, continuous conversational speech samples may require more time to obtain and possibly to transcribe. The central issue here is the benefit in validity in relation to the cost in efficiency. As reviewed in our paper, a large sector of the literature on articulation assessment seems to be concerned with efficiency issues, seemingly disproportionate to the emphasis on efficiency in other communication disorders and in other clinical disciplines. Although maximum efficiency is a

dividend in research and clinical decision making, researchers and clinicians will readily agree that efficiency is never a sufficient rationale for using less than the most valid available assessment procedure.

Validity. We have the most difficulty with the several places in this letter that suggest it is sufficient for data in child phonology research to be approximately accurate. We discern this perspective in statements such as the following: (a) "First of all one might quibble over whether [statistically significant differences on] 8 out of 23 consonants (35%) is a major or minor difference"; (b) "On the contrary, their data indicate that articulation tests can provide roughly comparable data in a more efficient manner . . ."; and (c) (with reference to the substitution data obtained from the articulation test) "One could argue that these are more useful data, since substitutions show directly how children are using particular sounds within their linguistic system." Each of these statements suggests a perspective on measurement that we do not share. Relative to (a), we have difficulty dismissing the potential importance to research and clinical practice of making even one error in the interpretation of findings that could be shown to be associated with mode of elicitation (see extended examples in Shriberg, Gruber, and Kwiatkowski, in press, and Shriberg, Kwiatkowski, and Gruber, in press). Relative to (b) we have difficulty with the measurement concept of "roughly comparable." And relative to (c) we have difficulty with the suggestion that statistically significant differences can be dismissed if they are not consistent with one's particular view of phonology.

In summary, we stand by the statistical findings indicating that citation form testing yields neither typical nor optimal measures of speech performance and our conclusion that such data continue to be a major source of variance in cross-laboratory research in child phonology.

Judith A. Morrison
University of Redlands
Redlands, CA
Lawrence D. Shriberg
University of Wisconsin-Madison

References

- Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders*, 37, 55–63.
- 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.
- Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (in press). Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research*.
- Shriberg, L. D., Kwiatkowski, J., & Gruber, F. A. (in press). Developmental phonological disorders II: Short-term speech-sound normalization. *Journal of Speech and Hearing Research*.
- 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, 779–798.

Received March 21, 1994

Accepted April 6, 1994

Response to Ingram Letter

Judith A. Morrison, and Lawrence D. Shriberg
J Speech Hear Res 1994;37:936-937

This information is current as of June 28, 2012

This article, along with updated information and services, is located on the World Wide Web at:
<http://jslhr.asha.org>

