Steps in Designing and Implementing an Innovative Assessment Instrument

Harry N. Seymour, Ph.D.,[1,2] and Barbara Zurer Pearson, Ph.D.[2]

ABSTRACT

Preliminary research for innovative assessments valid for both African American English- and Mainstream American English-speaking children suggested a process consisting of two separate tests: (1) a screening test, and (2) a comprehensive test of Syntax, Pragmatics, Semantics, and Phonology. Language probes were designed to accomplish the functions of dialect identification, using highly contrastive features between the dialects, and diagnosis of disorder, which uses noncontrastive elements. The resulting assessment/proposal, which has undergone extensive experimental field testing, differs from existing tests at the level of its individual items and in the process of test construction as a whole.

KEYWORDS: Dialect identification, diagnostic risk, language domains, dialect-neutral, dialect sensitive

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) unique aspects of the proposed dialect-sensitive test's design, development, and goals; and (2) how the proposed language probes avoid the typical pitfalls of standardized test design that often lead to linguistic bias.

INDEPENDENT EVALUATION OF DIALECT AND DISORDER

A basic premise for making a dialect-sensitive language assessment is that a test for dialect status does not tell about impairment. Most children who do not speak Mainstream American English (MAE), especially those with African American English (AAE)-speaking language models in their home environments, generally turn out upon further evaluation to have age-appropriate language. Children who score in the MAE range, whether African American or not, may also turn out to be language-impaired. Research for innovative

Evaluating Language Variation: Distinguishing Dialect and Development from Disorder; Editors in Chief, Nancy Helm-Estabrooks, Sc.D., and Nan Bernstein Ratner, Ed.D.; Guest Editors, Harry N. Seymour, Ph.D., and Barbara Zurer Pearson, Ph.D. *Seminars in Speech and Language*, volume 25, number 1, 2004. Address for correspondence and reprint requests: Barbara Zurer Pearson, Ph.D., Department of Communication Disorders, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: bpearson@comdis.umass.edu. [1]Professor Emeritus, [2]Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts. 0734-0478,p;2004, 25,01,027,032,ftx,en;ssl00180x.

assessments being developed by Seymour, Roeper, and de Villiers suggested an assessment consisting of two separate tests: (1) a screening test, and (2) a comprehensive test of Syntax, Pragmatics, Semantics, and Phonology (FN).

Proposed Test Format

The screening test we have been piloting (see Acknowledgments) consists of two parts: one part tests *contrastive* features of AAE phonology and morphosyntax, whereas the second part has all *noncontrastive* items. We found that a combination of noncontrastive morphosyntax and *wh*-question items, along with nonword repetitions,[1] provides a reasonably quick and reliable diagnosis of risk for language disorder that is not biased against AAE speakers. Children scoring in high-risk categories on the screener, regardless of dialect status, can be given a more comprehensive test for a fuller diagnosis.

To be comprehensive, a language test would need to cover the four traditional domains of linguistic description: Syntax, Pragmatics, Semantics, and Phonology. The comprehensive test could be given on its own or as a follow-up to an indication of the need for further evaluation provided by the screening items. Figure 1 provides a schematic diagram of how the various parts of the test might relate to each other.

IDENTIFYING IMPAIRMENT ON THE COMPREHENSIVE LANGUAGE TEST

The criteria for identifying impairment on the comprehensive test take into consideration that only a small minority of children (~4% in the developmental milestone research) will score in the failing (or "weakness") range in all domains. In fact, 20% of our research sample failed one domain. What is important for evaluating those children is how they performed in the other areas. Many children who fail one domain show normal or above-average performance in the others and so give no cause for concern. However, if the child fails two domains or is low average or below in two domains beyond the failing one, there is cause to recommend intervention.

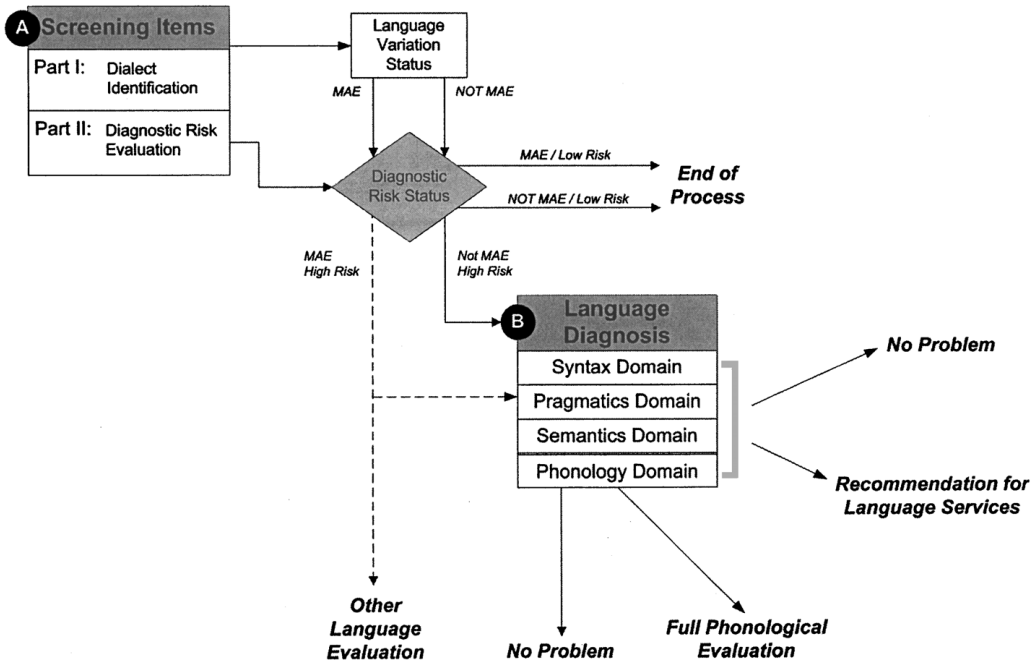

Figure 1 Relationship between elements of proposed assessments. (A) and (B) represent potential starting points for the assessment. MAE, Mainstream American English.

Special Consideration for Phonology Domain

The Phonology domain in the dialect-sensitive language test is an exception to this principle. It can count as a domain in adding up the child's strengths and weaknesses to evaluate the overall language performance. If a child failed Syntax and Phonology, the recommendation for remediation would be the same as if the child had missed two of the other domains, for example Syntax and Semantics, or Semantics and Pragmatics.

However, the Phonology domain can also stand alone for the evaluation of speech problems. About half of the children in the research described by Pearson[2] who had low Phonology scores also exhibited a language problem, but about half did not. With the dialect-sensitive Phonology as envisioned for our project, a low average or weakness result in that single domain would be sufficient to suspect a speech problem. Regardless of dialect background, the child should be given a full phonological evaluation. As mentioned above, if the child scored weakness in Phonology and one of the other domains, it would support a recommendation for language services as well as the speech evaluation.

UNIQUENESS OF THE PROPOSED TEST PROCEDURES

The dialect-sensitive assessment described in this issue is different from already existing tests in its design, development, and goals. The innovation is not in the nature of the stimulus items, which can be presented in as standard a format as possible. The dialect-sensitive probes that we propose require no special technology, nor any special knowledge of AAE (although it is always good practice to be as informed as possible about the characteristics of the children in one's care). Many of the items themselves may appear familiar, but closer consideration shows that they are not. The short narratives, passives, and many of the morphosyntax prompts to be described in this issue look like similar items on other tests. On the proposed assessment, however, their scoring and interpretation are unusual. Short narratives, for example, do not require recording and the child's story is evaluated not for its organization as is most common, but primarily for what it reveals about the child's language for mental events. The passive items are short and use common activities, but they represent a series of items of graduated difficulty according to the child's ability to understand implicit information about reported events as opposed to what is explicitly stated in the prompts. Elements of morphosyntax are elicited, as in other tests, but their purpose here is principally dialect identification, and they play only a small part in the diagnosis of impairment.

Other items may appear somewhat strange. Many observers have commented that the barriers questions in *wh*-questions (in the proposed Syntax probe) are too difficult for young children or that the double *wh*-questions are not the kind of questions that children typically hear. They are not easy, but the extensive milestone research has assured us that most typically developing children—mainstream and African American—can understand these questions and respond appropriately. Language-impaired children, by contrast, do find them hard, and do not demonstrate understanding of them until much later, if at all. Likewise, some of the fast mapping questions (in Semantics) often make adults pause—before getting them right. It is rare for an adult English speaker to miss more than an occasional one of the items. Children, too, are remarkably successful with them, although of course few of the younger children get the full syntax, for example, of novel complement items, the most complex of that item type. Nonetheless, even young children can demonstrate their understanding of how sentence grammar helps fix meaning in the easier transitive sentences, or in using the more direct relationships, such as subject and object, within the harder constructions.[3]

The proposed assessment's unique design originates in its commitment to being dialect-neutral. It was not written first and then tested with different populations and "tweaked" to accommodate group differences. It was put together exclusively with elements valid from a theoretical point of view for both AAE- and MAE-speaking children and which demonstrated empirically that they indeed were valid.

Except for the phonology and morphosyntax items, which make up the language variation identifiers, all elements of the two proposed measures have demonstrated that they are noncontrastive between dialects.

Overcoming Linguistic Bias

The dialect neutrality of the language probes is accomplished in one of several ways, both in the individual items and in the assessment's construction as a whole. The most basic way they avoid differences between dialects is to use fundamental structures that follow principles of universal grammar and are the same in different dialects of English. As described by Roeper[4] the ways *wh*-words in complex sentences can move across clauses appear to follow universal logical principles[5] and are essentially the same across many very distinct languages and across dialects of English.[6,7] Work by Roeper, de Villiers, and others[8,9] has shown that children 3 years and older are generally sensitive to those rules. *Wh*-clauses, then, reinforce dialect neutrality. Similarly, requests for information from one speaker to another may be encoded in different language structures, but the basic task must be accomplished by speakers of all languages. Asking a child to recognize what information is missing and to ask for the right information, as in the proposed question-asking subdomain in pragmatics, is a task of basic communication that speakers of all dialects should be able to demonstrate. As long as the scoring of the items depends on whether the child used functional language to accomplish the task—and not to produce particular target forms in doing so—this too can tap a type of language universal.

In some subdomains, as in the narrative section of the proposed pragmatics probe, there is strong evidence in the literature that the cultural groups differ, even in important respects,[10–12] but the parts of the task we suggest scoring (reference contrast and theory of mind) do not differ across dialects. In areas such as verb contrasts in semantics, where uses of many vocabulary items might well be expected to differ in different communities, only the specific items that showed no difference between dialect groups were selected.

Finally, great pains were taken to ensure that the artwork for the proposed measures was inclusive from a multicultural point of view. There are no exotic animals, no holiday references, and the children depicted represent all the ethnicities in the United States (but with a preponderance of African American children). Few items include proper names, which are often culture-specific and can add to the memory load of items; most use pronouns or generic terms such as "this boy," "this girl," etc. Whenever possible, there was an occasional bit of humor.

The proposed assessments are also unique in the choice of the populations for their standardization samples. The primary participants in the experimental field testing were AAE speakers of working class background, precisely the group least well served by current tests. Only after the AAE speakers assured the authors that the scoring would be valid and informative for non-MAE speakers were analyses of the performance of an MAE-speaking comparison group performed to demonstrate that the tests would work equally well for them.

A final step to ensure that these tasks are appropriate instruments for AAE speakers will be the standardization on African American children and a means for making this type of assessment commercially available.[13] However, even before that is accomplished, its precursors have already demonstrated how underlying linguistic principles are subject to disorders and can be a useful basis for the assessment process for both AAE and MAE speakers.

ACKNOWLEDGMENTS

This work was funded in part by National Institutes of Health (NIDCD) under Contract # N01 DC8-2104 and Grant # R01 DC 02172-04 to Harry Seymour, Principal Investigator, at the University of Massachusetts Amherst, with Thomas Roeper and Jill de Villiers at the University of Massachusetts and Smith College, as co-investigators. It was accomplished in conjunction with The Psychological Corporation of Harcourt Assessment, Inc., San Antonio, TX.

The tests that are the products of this research collaboration are the *Diagnostic Evaluation of Language Variation (DELV)*

assessments, the *DELV Screening Test, DELV Criterion-Referenced* edition, and the *DELV Norm-Referenced* edition. The phrase "evaluating language variation" refers generally to the assessment processes discussed in this issue. The term *DELV* is the name trademarked by The Psychological Corporation of Harcourt Assessments, Inc., and refers to the specific tests that are the outcome of the extensive research described in this article. The specific tests are referred to as the *DELV-ST*, or "screener," or the *DELV-CR, DELV-NR*, or the "full diagnostic test," as appropriate. Questions about the principles underlying the tests can be referred to the authors of this issue (Seymour, Roeper, de Villiers, de Villiers, Pearson, and Ciolli). Questions about the tests themselves should be addressed to the Project Leader at The Psychological Corporation of Harcourt Assessment, Inc.; Lois Ciolli, Senior Research Director.

REFERENCES

1. Campbell T, Dollaghan C, Needleman H, Janosky J. Reducing bias in language assessment: processing-dependent measures. J Speech Lang Hear Res 1997;40:519–525
2. Pearson BZ. Theoretical and empirical bases for dialect-neutral language assessment: the contribution of theoretical and applied linguistics to communication disorders. Semin Speech Lang 2004; 25:xxx–xxx
3. de Villiers JG. Cultural and linguistic fairness in the assessment of semantics. Semin Speech Lang 2004;25:xxx–xxx
4. Roeper T. Diagnosing language variations: underlying principles for syntactic assessment. Semin Speech Lang 2004;25:xxx–xxx
5. Chomsky N. Barriers. Cambridge, MA: MIT Press; 1986
6. Bland L, de Villiers J, Roeper T, Champion T, Seymour HN. Wh-movement in children of divergent language backgrounds. Presented at: the American Speech-Language-Hearing Association Annual Meeting. San Antonio, TX; 1992
7. Weissenborn J, Roeper T, de Villiers JG. The acquisition of WH-movement in German and French. In: Maxfield TL, Plunkett B, eds. Papers in the Acquisition of WH. Proceedings of the University of Massachusetts Roundtable. Amherst, MA: UMOP Special Edition; 1990: 43–73
8. de Villiers JG, Roeper T, Vainikka A. The acquisition of long-distance rules. In: Frazier L, de Villiers JG, eds. Language Processing and Language Acquisition. Boston: Kluwer Academic; 1990: 257–297
9. de Villiers JG, Roeper T. Relative clauses are barriers to wh- movement for young children. J Child Lang 1995;22:389–404
10. Gee JP. Two styles of narrative construction and their linguistic and educational implications. J Educ 1989;171:97–115
11. Michaels S. Listening and responding: hearing the logic in children's classroom narratives. Theory Pract 1984;23:218–224
12. Champion T. Understanding Storytelling among African American Children: A Journey from Africa to America. Mahwah NJ: Lawrence Erlbaum; 2002
13. Seymour HN, Roeper T, de Villiers JG. Diagnostic Evaluation of Language Variation (DELV) Norm-Referenced Edition. San Antonio TX: The Psychological Corporation; 2005 (projected)