University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Publication of the DELV tests and beyond

NIH Working Groups on African American English (AAE)

2004

Evaluating language variation: Distinguishing development and dialect from disorder. Special issue

Harry N. Seymour
University of Massachusetts Amherst

Barbara Zurer Pearson
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/aae_delv

Seymour, Harry N. and Pearson, Barbara Zurer, "Evaluating language variation: Distinguishing development and dialect from disorder. Special issue" (2004). *Seminars in Speech and Language*. 9.
Retrieved from https://scholarworks.umass.edu/aae_delv/9

This Article is brought to you for free and open access by the NIH Working Groups on African American English (AAE) at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Publication of the DELV tests and beyond by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

online www.thieme-connect.com

Seminars in Speech and Language

VOLUME 25, NUMBER 1 2004

Evaluating Language Variation: Distinguishing Dialect and Development from Disorder

Harry N. Seymour, Ph.D., and Barbara Zurer Pearson, Ph.D.
Guest Editors

CONTENTS

1 Preface
Nan Bernstein Ratner, Ed.D.

SECTION I: INTRODUCTION AND BACKGROUND

3 The Challenge of Language Assessment for African American English-Speaking Children: A Historical Perspective
Harry N. Seymour, Ph.D.

13 Theoretical and Empirical Bases for Dialect-Neutral Language Assessment: Contributions from Theoretical and Applied Linguistics to Communication Disorders
Barbara Zurer Pearson, Ph.D.

27 Steps in Designing and Implementing an Innovative Assessment Instrument
Harry N. Seymour, Ph.D., and Barbara Zurer Pearson, Ph.D.

33 Dialect Identification versus Evaluation of Risk in Language Screening
Lois Ciolli, M.A., C.C.C.-S.L.P., and Harry N. Seymour, Ph.D.

THIEME
New York • Stuttgart
Copyright © 2004 by Thieme Medical Publishers, Inc.
333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662.

SECTION II: KEY CONCEPTS AND EXAMPLES FOR THE DOMAINS OF THE PROPOSED TESTS

41 Diagnosing Language Variations: Underlying Principles for Syntactic Assessment
Thomas Roeper, Ph.D.

57 Assessing Pragmatic Skills in Elicited Production
Peter de Villiers, Ph.D.

73 Cultural and Linguistic Fairness in the Assessment of Semantics
Jill G. de Villiers, Ph.D.

91 A Noncontrastive Model for Assessment of Phonology
Harry N. Seymour, Ph.D.

SECTION III: PRACTICAL ISSUES AND GOALS FOR FUTURE RESEARCH

101 Distinguishing Dialect and Development from Disorder: Case Studies
Barbara Zurer Pearson, Ph.D., and Lois Ciolli, M.A., C.C.C.-S.L.P.

113 Conclusions, Future Directions, and Implications for Remediation
Harry N. Seymour, Ph.D., Thomas Roeper, Ph.D., and Jill G. de Villiers, Ph.D.

C-1 Continuing Education Self-Study Program

The Challenge of Language Assessment for African American English-Speaking Children: A Historical Perspective

Harry N. Seymour, Ph.D.[1]

ABSTRACT

The diagnostic problem of how validly to assess the language of children who speak dialects different from Mainstream American English (MAE) has challenged the field of communication disorders for several decades. The key to its solution is to recognize differences due to dialect or development and remove them from the initial diagnosis of a disorder. A new approach to the puzzle, implemented jointly by University of Massachusetts scholars and the Psychological Corporation (TPC), takes two directions: (1) it provides new normative data on African American English (AAE) development, and (2) it proposes a level of analysis deeper than dialect for the discovery of alternate markers of a disorder. We present three objectives for a language assessment instrument designed to solve this longstanding problem: (1) to answer the problem/no problem question for a given child; (2) to provide explanatory data about the nature of the problem; and (3) to achieve objectives 1 and 2 in a way that is culturally and linguistically fair to both speakers of MAE and speakers of other dialects of English such as AAE.

KEYWORDS: Sources of language variation, Mainstream American English (MAE), African American English (AAE), cultural and linguistic bias, contrastive and noncontrastive language features, ASHA's position paper on social dialects

Learning Outcomes: As a result of this activity, the participant will be able to describe (1) the three aspects of language that must be addressed when attempting to validly assess the language of a child who is speaking African American English (AAE), and (2) the clinical problems presented when attempting to validly assess the language of a child who is speaking AAE.

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: Harry N. Seymour, Ph.D., Department of Communication Disorders, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: hseymour@comdis.umass.edu. [1]Professor Emeritus, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01,003,012,ftx,en;ssl00178x.

DEVELOPMENT, DIALECT, OR DISORDER?

Story A: *The big brother has the train. And he held it up high so he couldn't get it and then he hide... put it under the bed. And then his little brother, he looked under the bed when his brother was eating his sandwich. And then he put it in his toy box. Then the big brother was thinking about the train, and he looked under the bed, but it wasn't there. (Ages 6;1)*

Story B: *He's not giving the train back to him. He put it under his bed. He eats a sandwich. He gets his train under his bed. He comes in there. He puts it in his toy box. Then he comes lookin' for it. Then he looks under the bed. (Ages 4;8)*

Story C: *Once upon a time a little boy wanted to play with the train. And he tryna get the train, but his mean brother holdin' it up high. So the big boy put it under his bed. When the big brother in the kitchen eatin' his sandwich, the little boy take he train and he put it in his toy box. Then the big boy came back and he thinkin' of the train. And he look under the bed but he don't find nothin'. (Ages 6;3)*

Story D: *The big guy... the wittle brother wants to play with the choo-choo train. The big guy hiding it under his bed. The wittle guy... the wittle brother gotten it and the wittle brother putting it in his toy box. Then the brother came in the door. Then he came look for the toy. Then look under his bed. (Ages 6;7)*

The narratives presented above show four very different children. Three of them differ in expected ways. One differs in ways characteristic of a child with a language disorder.

To identify which narrator shows evidence of a language disorder, three aspects of child language variation must be addressed. The evaluation of a disorder must first take into account the child's dialect status and developmental status. In determining which of these narrators has a disorder and is in need of language services, the speech-language pathologist (SLP) must distinguish which elements derive from the child's stage of development, which elements reveal the child's dialect, and which elements indicate disorder.

Making the distinctions among development, dialect, and disorder for the above narratives can represent a difficult diagnostic puzzle. This puzzle is relatively simple when only mainstream American English (MAE) is involved. There is an extensive literature describing the milestones of development and how children learn to use language in appropriately formed narratives as they grow and mature, so the developmental distinction between Story A and Story B is clear. Story A is produced by a 6-year-old child and Story B by a 4-year-old child. If a 6-year-old child produced the story in B, it might be cause for concern, but B is perfectly normal for a 4-year-old child.

On the other hand, Story D is produced by an MAE-speaking 6-year-old child. The syntactic forms are inconsistent and unpredictable; the narrative focuses only on the actions in the pictures and lacks the cohesive devices expected of a child of that age. Again, the diagnostic puzzle is not too difficult to resolve for an MAE-speaking child with a language disorder given the indicators of delay in Story D.

However, in the case of Story C the situation becomes more complicated in that the narrator is not a speaker of MAE and uses stigmatized language forms, which may distract the listener from recognizing the basic quality of the story; and because of various omissions of morphological inflections, this child could be confused with a child who has a language disorder. In fact, Story C is told by an African American English (AAE) speaker who, like the speaker in Story A, exhibits above average use of narrative markers, such as time clauses (e.g., "when") and sophisticated knowledge of the language of thought (e.g., "he think about...."). To diagnose the speaker of Story C as having a disorder would be a serious but not an uncommon mistake if AAE status were not factored into the diagnosis.

THE DIAGNOSTIC CHALLENGE

The diagnostic problem associated with the above narratives captures the essence of a clinical conundrum that has challenged communication disorders for several decades. This challenge is how validly to assess children who speak dialects so different from MAE that it is

difficult to separate them from children with a language disorder. In this issue, solutions to this clinical problem are presented so that the AAE status of Case C is recognized for what it is, a dialectal variation, and not confused for the disordered status of Case D.

Making this distinction, that is, distinguishing the difference between typically developing MAE- and AAE-speaking children from their language-impaired peers, is the objective of this issue. The objective is achieved through an in-depth discussion of the development, design, and research findings of a project to develop an innovative assessment test for a non-MAE population. This research and the resulting test construction were focused on how best to isolate dialect factors and remove them from the initial diagnosis of risk. The proposed battery exemplifies the separation of two functions: dialect identification and diagnosis of a disorder. These tests and the process they derive from are used here to demonstrate how children, including AAE speakers, can be evaluated fairly and without linguistic or cultural bias.

The dialect-neutral language test described in this issue was developed by Harry Seymour, Tom Roeper, and Jill and Peter de Villiers at the University of Massachusetts Amherst in conjunction with The Psychological Corporation. The extensive research foundation of the test was sponsored by a grant and then a contract from the National Institutes of Health-National Institute on Deafness and Other Communication Disorders to develop a language assessment instrument that would be appropriate and unbiased for African American children who speak AAE. It has taken more than 10 years and the work is not yet finished. Preliminary versions of a screener and a comprehensive language test were published in 2003. A norm-referenced version with a significant sampling of African American children is being developed now for publication in 2005.

Clinical Implications of Dialect Differences

In general, SLPs rely heavily on a single dialect standard (i.e., MAE) as the referent of acceptability when assessing the language of children. Although MAE is an abstract notion and is not a specific language entity, it is nevertheless perceived to be the variety of English most used in the conduct of commerce and is fostered in the schools as most acceptable. Moreover, the notion of a single MAE is a misnomer in that MAE can vary from one region of the country to another. For example, the MAE spoken in Boston sounds different from the MAE in Georgia.

Indeed, President John F. Kennedy, who dropped his /r/, as in "pak the ka", sounded very different from President Jimmy Carter, whose accent was distinctly southern. Of course these differences are superficial variations of English, and are considered to be MAE even though they represent two very distant communities, a northern city and the rural south. Despite such variations in English speech patterns across the country, there exists a common core of language features that defines who is and is not a speaker of MAE. Hence, differences among MAE speakers are relatively minor and a standard archetype is generally applied by SLPs despite regional variations.

Most variations among dialects of English are relatively superficial, representing simple contrasts from the archetype standard, and thus, present little difficulty to SLPs in accommodating those variations in the assessment process. However, there are a few dialects such as AAE in which there are more profound contrasts from MAE and the assessment process becomes more complicated.

AAE has been described by sociolinguists over the last several decades as a dialect primarily spoken by African Americans. It is variously referred to as Black English, Ebonics, African American English Vernacular, among other names. (The preferred term in this article is AAE, which is commonly used in academic circles.) AAE, like MAE, is but one of many varieties of English. However, unlike MAE, which cuts across geographic, racial, and ethnic boundaries, AAE is characterized by a commonality of speech spoken primarily by African Americans, but not by all of them. AAE is less geographically defined than other dialects of English, though there are some differences by geographic region; rather it has emerged as a commonality of speech and grammar of a culturally defined group. In addition, children or adults of other races who have strong cultural

identification or primary social interaction with African Americans may speak AAE too. Thus, AAE may be defined in terms of the features that distinguish a pattern of grammar (morphology, semantics, syntax, pragmatics, and phonology) in the speech used by culturally identified African Americans.

The Challenge of AAE Diagnosis

The clinical problems presented by AAE are 2-fold: (1) AAE features appear similar to patterns of language disorders, and (2) there is a paucity of developmental research on AAE. With respect to the first issue, AAE is heavily characterized by optional use of certain linguistic structures, particularly morphological inflections. The term "optional" refers to a speaker's use of a particular linguistic structure at some times and not others. For example, AAE speakers may or may not produce a present tense sentence without a copula verb (i.e., "a zero-copula," *He is tall → He tall*), zero third-person present tense agreement (*He walks → He walk*), and zero past *-ed* (*He played yesterday → He play yesterday*). Because, it is not uncommon for a language-disordered child to delete copula "is," third person /*-s*/, and past /*-ed*/, it may be difficult, in a diagnostic context, to differentiate such disordered patterns from the typical patterns spoken in AAE.

This differentiation problem is further exacerbated by issue two, the limited information about the course of development of such patterns as copula, third-person /*-s*/, and past *-ed*. As AAE children's language matures during the acquisition stage, it remains unclear when these and other AAE patterns are mastered and the forms they take. In MAE, there are specific normative milestones that indicate when children of various ages acquire mastery of morphological inflections. These milestones constitute acquisitional benchmarks for SLPs to follow in identifying children who fail to achieve them. A similar set of benchmarks is lacking for AAE-speaking children.

Despite the absence of a comparable dataset for AAE acquisition, undoubtedly there are strong similarities between the two dialects, given that both MAE and AAE are spoken by speakers of English. The important question about AAE acquisition is when particular AAE features are mastered. During the acquisitional stage, an informed position is that both AAE and MAE produce very common developmental patterns such as deletions of morphological inflections.[1–3] As both dialects mature in the process of acquisition, they diverge such that each adopts the adult patterns of its respective system. This means that many of the developmental patterns shared between the dialects become extinct in MAE, but appear to be retained in AAE. For example, at age 3 years, both dialects may produce an absent third-person /*-s*/ agreement marker; at age 5 years this feature no longer exists for the MAE speaking child, but remains in the AAE-learning child's speech at 5 years and older. Because of the optionality of features such as the third-person /*-s*/ agreement marker, the point at which a child's production shifts from immature status to adult status has not been determined.[4,5] At the same time, uniquely AAE elements, such as the use of an invariant form of "to be" to indicate habitual actions, are emerging in the AAE-learning child. Additional complexities such as these intersect with the third-person /*-s*/ agreement system in ways that have not been investigated. Thus, this acquisition puzzle remains a source of difficulty when assessing child AAE speakers.

HISTORICAL PERSPECTIVE IN ADDRESSING THE CLINICAL PROBLEM

A debate between John Michels and Orlando Taylor[6,7] during the 1968 American Speech-Language-Hearing Association (ASHA) annual convention brought forward these controversial issues about linguistic and cultural bias in the assessment and treatment of African American children. As a result of this debate and the issues it raised, a small group of African American scholars formed the ASHA Black Caucus. Through the efforts of this Caucus, ASHA was motivated to revise its curriculum requirements for clinical certification by broadening the recommended coursework to include the study of sociolinguistics and the topic of AAE.

Throughout the 1970s, the Black Caucus continued to push for fair and equitable testing practices and a multicultural perspective in ASHA. By 1978, the Caucus had evolved into the National Black Association for Speech-Language and Hearing, an organization that has been very effective in nurturing and disseminating scholarship among African American students, SLPs, and university faculty regarding the communicative styles of African Americans. In addition, the organization has played an important role in the adoption of ASHA's position paper on social dialects, which states "... no dialectal variety of English is a disorder or a pathological form of speech or language."[8,9]

ASHA's position paper on social dialects established that SLPs should not view AAE as a deficit. This position was helpful in providing a challenge to the deficit position, which argued that AAE was cognitively corrupt and deficient. ASHA's position paper contradicted this argument and gave support for an opposing position, that AAE was simply different and was a rule-governed dialect of English. Although a significant step forward in acknowledging the legitimacy of AAE, the position paper fell short of telling SLPs what to do about AAE in terms of assessment and treatment of communication disorders among AAE speakers.

The testing practices in communication disorders during the last several decades are such that reliance on standardized tests of language almost ensures that children who speak AAE are likely to be penalized for doing so. One form this penalty takes is a disproportionate representation of African American children in language services and special education programs throughout the country. This was so in 1970 and it prevails to this day.

Consider that across the United States, African American children are three times more likely to be diagnosed as mentally retarded than their white peers.[10] This disproportionality is also reflected in special education classifications.[11] African American children comprise ~15% of the school-age population, yet their numbers in special education for disabilities associated with language functioning constitute an over-representation in every category. All of these educational designations rely at least in part on testing done in the medium of MAE (Fig. 1).

One can reasonably argue that where there is over-representation, there are undoubtedly children who are in fact language disordered and who go undiagnosed. This would be under-representation, which, along with over-representation, reflects misdiagnosis and can be attributed in part to an assessment process that is biased against African American and other minority children. This bias can take several forms.[12] According to Wyatt[13] there can be situational bias, format bias, value bias,

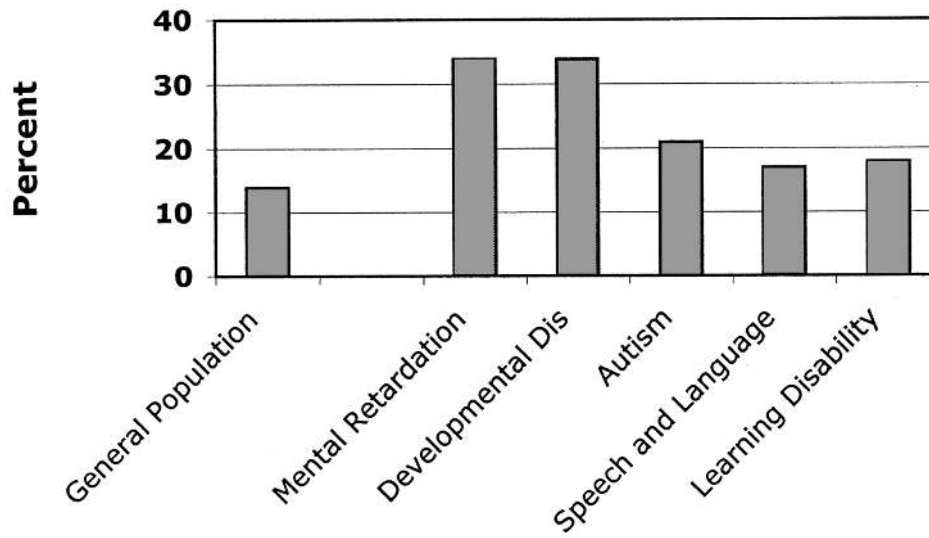


Figure 1 Percent of African American children in special education classifications (Dis, disorder).

and linguistic bias, all of which create a disconnect between the client and the testing context.

Of these various kinds of bias, perhaps the most difficult to address in the clinical setting is the linguistic and cultural bias. This difficulty stems from the importance and reliance on standardized testing. Because standardized tests typically are normed on racial and ethnic groups in accordance to a demographic distribution that matches the U.S. general population, there is a linguistic and cultural skewing toward the majority group; that is, middle-class white Americans. As a consequence, linguistic and cultural differences associated with dialects such as AAE are not adequately reflected in the normative distribution of the tests. Thus, AAE speaking children can be and are penalized for their dialect.

MULTICULTURAL MODELS

Language is clearly a major factor in the kinds of special education disabilities for which African American children are over-represented. This language factor is directly related to the mismatch between the target MAE standard on assessment tests and the AAE status of so many African American children. In recent years and due largely to a society now characterized by diverse ethnic, racial, and language communities, many if not most SLPs recognize that they can no longer apply a single linguistic and cultural model in assessing and treating language disorders in children. Serious consideration must be given to the language background from which children come in determining the kinds of assessment strategies and materials that are most appropriate.

With respect to AAE, these assessment strategies and materials must draw upon existing knowledge about AAE. Much of this knowledge derives from linguistic descriptions about AAE from the early work of sociolinguists during the 1960s and early 1970s.[14–21] This seminal work became foundational for later descriptive research and for those focusing on the clinical issues concerned with distinguishing language deficits from language differences characterized by AAE.[4,12,22–27]

Research on AAE has led to the proposal of several models for assessing AAE-speaking children. Some have suggested abandoning standardized tests for African American children altogether in favor of "nonstandardized" assessment methods.[23,28,29] Among these methods are language sampling analysis and criterion-referenced language probes. As useful approaches in general, both language sampling and language probes can be recommended specifically for AAE children because their naturalistic and dynamic process can be free of bias compared with the relatively restrictive and artificial testing contexts of standardized tests. However, such approaches carry the disadvantages of being time intensive, possibly less reliable, and having inadequate normative data on AAE.

Because of the limited normative data on AAE and remaining questions about the system of AAE, some scholars have directed their attention at those linguistic structures that are common to both AAE and MAE. This emphasis avoids AAE features and also avoids the clinical problem regarding the status of optional AAE structures. One such approach is the Minimal Competency Core (MCC) proposed by Stockman.[28] The MCC identifies a scale of obligatory language patterns expected of all typically developing children regardless of their dialect status. This scale constitutes a criterion-referenced measure representing the lowest end of a competency scale against which acceptable language performance can be measured. In addition, Craig and Washington[30] focus on complex sentence constructions common to both AAE and MAE. This approach avoids zero morphosyntax patterns so typical of AAE. There is also support for avoiding AAE features in the diagnostic process in the work of Seymour, Bland-Stewart, and Green,[31] who showed that AAE features that contrast with MAE were less effective in identifying language disorders among African American children than features that were noncontrastive between AAE and MAE

This contrastive/noncontrastive analysis, first proposed by Seymour and Seymour in 1977,[4] introduced a diagnostic procedure that focuses entirely on those language structures that AAE and MAE have in common; that is, those that are noncontrastive. The underlying assumptions for this approach are (1) AAE and

MAE are more similar than they are different; (2) specific AAE features should be avoided because they represent patterns that appear similar to disordered features, and thus would be ambiguous in the diagnostic process; and (3) an impaired language system also will reflect itself in the similarities between AAE and MAE. The intention was that this noncontrastive emphasis would answer the problem/no problem question in diagnosis; that is, is there a problem or not? It was fully recognized that to determine the nature of the language problem, a complete diagnosis should then involve the child's full array of language strengths and weaknesses, which would, by necessity, involve both contrastive and noncontrastive language structures. The problem in implementing such a scheme and in the absence of standardized tests was that a clinician would have to employ time-consuming and technically challenging linguistic analysis associated with language sampling and language probes. As a consequence, too often the clinician has to rely on existing standardized tests as a default procedure to both answer the problem/no problem question and to determine the nature of the problem. The consequence is a perpetuation of inappropriate and biased practices for AAE-speaking children.

THE PROPOSED SOLUTION

At the heart of the clinical problem associated with dialects such as AAE is the issue of variation. SLPs must examine and interpret at least three important sources of variation when diagnosing language disorders in children. These are variation due to speech and language development, variation due to speech and language dialects, and variation due to speech and language disorder. These sources of variation account for how children's language may differ from the adult model, and adequate assessment must first deal with the evaluation of that language variation.

A test designed to determine sources of linguistic variation must fulfill three objectives: (1) to answer the problem/no problem question; (2) to provide explanatory and evidentiary data about the nature of the problem; and (3) to achieve objectives 1 and 2 in a way that is culturally and linguistically fair to both speakers of MAE and speakers of other dialects of English such as AAE. Our proposed battery (see Acknowledgments) meets these objectives by splitting the process in two: Step 1 starts with dialect identification, elicited with contrastive structures; Step 2 follows with dialect-neutral diagnosis of disorder. Once dialect identification has been accomplished in the first step of the process, the second part of the screening can concentrate on the diagnostic function. For this, diagnostic test items avoid superficial contrasts between dialects of English by focusing on structures that are noncontrastive and by drawing upon deep principles of language considered universal across dialects and even languages. Such items will be among the most difficult and challenging for children—although they are also among the most effective as language assessment tools.

Most tests of language address the problem/no problem question only and offer little to no explanation about the nature of the child's problem.[32] The reason for this limited focus is the fragmented and atheoretical properties of the items composing the tests. Typically, language test items are selected for their capacity to measure a discrete aspect of language and provide a developmental differentiation across ages. Consequently, the outcome measures are gross indices of a child's knowledge. Our proposed assessment, on the other hand, will be a process-driven comprehensive test of language composed of test items that are theoretically coherent within language domains. Hence, the outcome measures for children will yield rich information about the child's understanding of processes as opposed to discrete and unrelated bits of language.

If the proposed battery is an effective assessment regardless of the child's dialect and thus could be given to any child, one might question why it is necessary to evaluate Language Variation Status. However, in the case of AAE-speaking children, there are still compelling reasons to determine a child's variation status. First, it helps document which individuals within the African American community are mainstream speakers. Race alone is not

a sufficient indication. Among the African American subjects in our field-testing research, for example, 15% percent were identified as MAE speakers by the screener (and fewer than 20% of those children were recognized as MAE speakers by their teachers). A teacher's knowledge that a child speaks MAE or not can be useful in general educational programming, such as in planning reading instruction. Furthermore, the greatest value of documenting a child's dialect is to the nonimpaired AAE speakers. They cannot be considered impaired solely on the basis of the finding that they do not speak MAE. If their risk status is high, then those children need further evaluation to make a diagnosis of impairment, and that evaluation should not be made with a test normed on MAE speakers.

CONCLUSION

In brief, the authors represented in this issue have designed an assessment process that consists of three tests appropriate for both MAE and non-MAE speakers between the ages 4 and 9 years. This issue is designed to share our developmental process with researchers and clinicians, and to further the development of additional dialect-sensitive measures. The first measure is a short screener with two parts: one part is designed to identify Language Variation Status in terms of whether a child is an MAE speaker or not, and the second part is designed to identify children who may be at risk for a disorder (Diagnostic Risk Status). The follow-up test is longer and provides criterion-referenced cut-off scores for a comprehensive assessment of syntax, semantics, pragmatics, and phonology. (See Figure 1 in Seymour and Pearson, this issue.) A norm-referenced version, essentially identical in composition to the criterion-referenced version, is projected for 2005. The norm-referenced sample for the test will incorporate a comparison to an oversampling of African American children, and thus will be the only test of language with a significant African American representation. How such tests and future assessment measures make possible a valid and unbiased assessment of children's language is fully described in the articles to follow.

Editor's Note

The material in this issue pertains especially to AAE and its speakers. Preliminary research indicates that certain other dialects of English, such as Cajun English and Appalachian English, share many features with AAE, and in pilot testing those children performed like AAE speakers on the proposed tests. However, the utility of such tests for speakers of Spanish-influenced English has not been established. In principle, young Hispanics in the process of learning English are using an "interlanguage," which is only partly English. Therefore, there is no reason to expect that the deep syntactic principles found in the tests described here will be realized in the interlanguage(s) in the same way that they are in varieties of English. On the other hand, if the Hispanic child has mastered MAE sufficiently to score in the MAE range in Language Variation Status, there is no reason she or he cannot be evaluated using the these instruments. However, the wider range of Hispanic-background children will be better served by a test designed for children in their circumstance, like the one currently under development by Iglesias et al.[33]

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 Assessment, 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. Cole L. A developmental analysis of social dialect features in the spontaneous language of preschool Black children. Dissertation, Northwestern University. Evanston, IL; 1980
2. Kovac C. Children's acquisition of variable features. Dissertation. Georgetown University, Wash., DC; 1980
3. Blake I. Language development in working-class black children: An examination of form, content, and use [dissertation]. 1984
4. Seymour HN, Seymour CM. A therapeutic model for communicative disorders among Black English speaking children. J Speech Hear Dis 1977;42:247–256
5. Seymour HN, Seymour CM. Black English and standard American English contrasts in consonantal development for four and five-year-olds. J Speech Lang Hear Dis 1981;46:274–280
6. Michel J. The role of ASHA in social, political, and moral activities. ASHA 1969;11:219–220
7. Taylor O. Social and political involvement of the American Speech and Hearing Association. ASHA 1969;11:216–218
8. American Speech-Language-Hearing Association Position paper: Social dialects and implications of the position on social dialects. ASHA 1983; 25:23–27
9. American Speech-Language-Hearing Association. Technical Report: American English Dialects. ASHA 2003;(suppl 23): (In press)
10. Parrish T. Racial disparities in the identification, funding, and provision of special education. In: Losen DJ, Orfield G, eds. Racial Inequity in Special Education, The Civil Rights Project at Harvard University. Cambridge: Harvard Education Press; 2002
11. Losen DJ, Orfield G, eds. Racial Inequity in Special Education, The Civil Rights Project at Harvard University. Cambridge: Harvard Education Press; 2002
12. Taylor O, Payne K. Culturally valid testing: a proactive approach. Topics Lang Disord 1983;3:8–20
13. Wyatt TA. Assessing the communicative abilities of clients from diverse cultural and language backgrounds. In: Battle DE, ed. Communication Disorders in Multicultural Populations, 3rd ed. Boston: Butterworth-Heinemann; 2002:415–459
14. Shuy R, ed. Social dialects and language learning: Proceedings of the Bloomington, Indiana Conference, August 1964. Champaign, IL: National Council of Teachers of English; 1965
15. Stewart W. Urban Negro speech: Sociolinguistic factors affecting English teaching. In: Shuy R, ed. Social Dialects and Language Learning: Proceedings of the Bloomington, Indiana, Conference, 1964. Champaign, IL: National Council of Teachers of English; 1965:10–19
16. Baratz JC. Language and cognitive assessments of Negro children: assumptions and research needs. ASHA 1969;11:87–91
17. Wolfram W. A Sociolinguistic Description of Detroit Negro Speech. Washington DC: Center for Applied Linguistics; 1969
18. Labov W. The logic of non-standard English. In: Williams F, ed. Language & Poverty. Chicago: Markham; 1970:153–189
19. Dillard JL. Black English: Its History and Usage in the United States. New York: Random House; 1972
20. Kochman T, ed. Rappin and Stylin Out: Communication in Urban Black America. Chicago: University of Illinois Press; 1972
21. Smitherman G. Talkin and Testifyin: The Language of Black America. Boston: Houghton-Mifflin; 1977
22. Adler S. Dialectal differences: professional and clinical implications. J Speech Hear Disord 1971;36:90–100
23. Seymour HN, Miller-Jones D. Language and cognitive assessment of Black children. In: Lass N, ed. Speech and Language Advances in Basic Research and Practice. New York: Academic Press; 1981:203–263
24. Terrell S, Terrell F. Distinguishing linguistic difference from disorders. Topics Lang Disord 1983;3:1–7
25. Vaughn-Cooke F. Improving language assessment in minority children. ASHA 1983;25:29–34
26. Stockman IJ, Vaughn-Cooke F. Implications of semantic category research for language assessment of nonstandard speakers. Top Lang Disord 1986; 6:15–25
27. Craig HK, Washington JA. An assessment battery for identifying language impairments in African American children. J Speech Lang Hear Res 2000; 43:366–379

28. Stockman I. The promise and pitfalls of language sample analysis as an assessment tool for linguistic minority children. Lang Speech Hear Serv Sch 1996;27:355–366
29. Leonard L, Weiss A. Application of nonstandardized assessment procedures to diverse linguistic populations. Top Lang Disord 1983;3:35–45
30. Craig HK, Washington JA. The complex syntax skills of poor, urban, African American preschoolers at school entry. Lang Speech Hear Serv Sch 1994;25:181–190
31. Seymour HN, Bland-Stewart L, Green L. Difference versus deficit in child African American English. Lang Speech Hear Serv Sch 1998;29: 96–108
32. Muma JR. Language Handbook: Concepts, Assessment, Intervention. Englewood Cliffs, NJ: Prentice Hall; 1978
33. Iglesias A, Peña E, Gutierrez-Clellen V, Bedore L, Goldstein B. Development of a language test for bilingual Spanish-English speaking children. (The Bilingual English/Spanish Assessment, BESA). Paper presented at: the American Speech-Language-Hearing Association Convention, San Francisco, CA, Nov. 18–21, 1999
34. The Council for Exceptional Children and the National Alliance of Black School Educators Addressing Over-Representation of African American Students in Special Education. Washington, DC; 2002:6

Theoretical and Empirical Bases for Dialect-Neutral Language Assessment: Contributions from Theoretical and Applied Linguistics to Communication Disorders

Barbara Zurer Pearson, Ph.D.[1]

ABSTRACT

Three avenues of theoretical research provide insights for discovering abstract properties of language that are subject to disorder and amenable to assessment: (1) the study of universal grammar and its acquisition; (2) descriptions of African American English (AAE) Syntax, Semantics, and Phonology within theoretical linguistics; and (3) the study of specific language impairment (SLI) cross-linguistically. Abstract linguistic concepts were translated into a set of assessment protocols that were used to establish normative data on language acquisition (developmental milestones) in typically developing AAE children ages 4 to 9 years. Testing AAE-speaking language impaired (LI) children and both typically developing (TD) and LI Mainstream American English (MAE)-learning children on these same measures provided the data to select assessments for which (1) TD MAE and AAE children performed the same, and (2) TD performance was reliably different from LI performance in both dialect groups.

KEYWORDS: Universal grammar, unique structures of AAE syntax, cross-linguistic study of LI, developmental milestones

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) the theoretical bases for a dialect-sensitive language screener and a dialect-neutral language assessment, and (2) the goals considered to be essential in the selection of the items for the proposed screener and diagnostic language assessments.

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]Research Associate, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01, 013,026,ftx,en;ssl00179x.

THE THEORETICAL FRAMEWORK

The need for a linguistically and culturally fair language test in communication disorders has been recognized since the 1960s, but until now there have been no standardized speech and language instruments specifically designed and constructed for learners of dialects other than Mainstream American English (MAE). In the interval, research in several fields has contributed to the conceptual basis for such a test. These research traditions are (1) the study of universal grammar and its acquisition within theoretical linguistics; (2) descriptions of African American English (AAE) Syntax, Semantics, Pragmatics, and Phonology within theoretical linguistics; and (3) the study of specific language impairment (SLI) cross-linguistically. Seymour, Roeper, de Villiers, and de Villiers (this issue) used insights from all three fields to create a practical instrument that takes into consideration typical development across dialects, and also the abstract properties of grammars to diagnose language disorder.

As described by Seymour in the introduction to this issue, following the social revolutions of the 1960s, sociolinguists gave us two important legacies. The first is an appreciation of the rule-governed nature of AAE, demonstrating that it is not a defective version of Mainstream English, but a logical and systematic language variation in its own right.[1] Sociolinguists of the 1960s and 1970s also devised a useful inventory of points of contrast between MAE and AAE,[2] which has been the basis of significant research in the intervening time.[3–5]

Universal Grammar

During the same period, abstract theoretical linguistics was elaborating a framework focused on a more abstract level of grammatical description. Linguists such as Chomsky[6,7] and colleagues posited a system of universal grammar within which all languages share a common underlying set of principles, but differ by how settings of certain parameters, or "switches," are set. This "principles and parameters" research provided a mechanism for understanding how small differences in the fundamental structures of languages would have effects in many different areas of the surface structure of languages. For example, if a language had noun phrases with the noun at the end of the phrase (such as Japanese), it would affect the typical word order of sentences in the language, the types and position of relationship words, such as prepositions, and the structure of relative and other adjunct clauses in the language. There are many other examples of the integral relatedness of parts of grammar within contemporary linguistics, and these informed the kinds of language probes described in this issue.

In the abstract study of phonology, likewise, Optimality Theory[8,9] provides a mechanism for us to view the greater underlying commonality between languages and dialects, while at the same time appreciating their surface differences. In Optimality Theory, researchers posit a restricted set of rules shared by all languages. Different languages "weight" the universal rules differently—that is, apply them in different orders or with different priorities—to give rise to many different outputs observed in different languages and dialects.[10,11]

With these frameworks as a background, the authors of this issue were motivated to look for underlying language principles as the foundation for the test, and as ways to diagnose disorder that would not confuse its signs with surface differences due to dialect. One such area, for example, is in the fundamental properties of movement rules and how movement is blocked, exemplified by *wh*- questions (see Roeper[12]). The theory behind syntactic barrier items, found in the discussion of syntax evaluation, derives from Ross's work on "islands,"[13] later elaborated by Chomsky in *Barriers*.[14,15] These principles were applied to child language acquisition theory by Roeper and de Villiers and their colleagues.[16,17]

Research on the Abstract Structure of AAE

Our work has also profited from greater exploration of AAE within the framework of abstract linguistics. *Wh*-barriers, for example, were investigated in an AAE-learning population,[18] and typically developing AAE-learning children were shown to respect the universal principles involved, much the same as children learning MAE or French or Greek, for

example.[19] In 1998, Coles[20] extended that work to provide new insights into the mechanisms governing negative concord, another AAE property, whereas Green[21,22] investigated the structural description of the AAE verb phrase. Green's work provides a possible explanation for why AAE verbs would not receive person marking (as in third-person *-s*) in the same way as MAE verbs. As Green points out, the AAE verbal system has a rich system of "aspectual" elements that have no counterpart in the grammar of MAE. ("Aspect" refers to other means aside from tense to describe the time element in an action—whether it was a continuing action, or a repeated action, one that happened all at once, or one that lasted a long time; e.g., "the leaf fell" versus "the leaf was falling.") For example, AAE has special meanings for verb auxiliaries, such as invariant *be* for habitual actions ("they [always] be running"), *done* for perfective, *been done* for remote past—all shades of meaning that require adverbs to express them in MAE. Green posits an additional branch in the AAE structural tree. As in Russian, which also has a richly marked system of aspect in its grammar, the aspect particles, which appear superficially the same as MAE tense markings, do not interact with the agreement system. If Green is right, this would give a deep structure reason for the lack of person agreement in the AAE verb phrase.

Other theoretically informed studies of AAE have been carried out by Terry,[23] Jackson,[24] Wyatt,[25] Dayton,[26] and others. Although not all of these findings have found direct translation into assessment probes, they emphasize the regularities in certain features of AAE, which were often considered instances of imperfect learning of MAE structures. In this framework, rather than being imperfect, the use of these same features can be viewed as learning an alternative grammar with properties as rich and complex as any other natural language.

Cross-Linguistic Study of SLI

Another important avenue of research is found in the recent interest in language impairment (LI) in languages other than English. The study of LI and SLI in different languages adds to our understanding of the deeper principles of language involved in such impairment. For example, much research on LI in English-learning children has focused on the acquisition of morphosyntax, and several theories attributed the difficulty to a problem with, for example, unstressed words and parts of words.[27] More recently, inspired by modern linguistic analyses of the functional categories, Rice and Wexler[28] have pointed to a particular difficulty in SLI in the area of tense. One of the recommendations of this line of work is to take a measure of a "finite verb morphology composite."[29] This measure represents the overall percentage with which children use morphemes such as past *-ed*, present third person singular *-s*, and copula and auxiliary *be* forms.[30,31] Leonard concludes, "For many children with SLI, the use of this collection of morphemes is unusually weak."

However, researchers studying children learning languages with more developed systems of morphosyntactic inflections report little difficulty with tense marking.[32] Instead, other difficulties come to light; for example, in agreement, as observed by Clahsen[33] or the incorporation of pronoun forms (clitics) into the verb phrase, as in French.[34,35] The problems with tense forms observed in English LI may be less a question of the morphology of tense, but rather a consequence of a more general problem in the elaboration of the verb phrase as a whole, which takes different forms depending on the computations required by the structural properties of the different languages[36]

In the case of AAE, even normally developing children measured by a standard such as the "finite verb composite" of MAE would be identified as having serious linguistic problems, because AAE is different in each one of these forms. As researchers, we sought to avoid confusing dialect variation and disorder and to capture the sophisticated abstract knowledge that all typically developing children learn without being taught, but which create problems for children with language impairment. We do not yet know why these problems might arise—whether it is, as has been suggested, a question of incomplete grammars, or processing difficulties, or missed parameters. Perhaps in SLI, the normal components of language

that feed growth to one another get into misalignment through delay in some area. Each of these, or yet another possibility, may be at work for different subsets of children with disorders. The theoretical basis for SLI is still an issue with many unresolved questions. Our final set of suggested tasks attempts to delve more deeply into areas where the child could have fundamental problems.

THE EMPIRICAL RESEARCH PROGRAM

Step 1: Developing Items

One way to go about translating theoretical background into an empirical research program that would identify dialect and culture-free assessment items was to mine protocols from acquisition experiments within Syntax, Semantics, Pragmatics, and Phonology for candidate items. Several of de Villiers and Roeper's experiments[17,37] provided formats for our *wh*-question probe to identify language impairment. Work in the acquisition of passive by Roeper,[38] question-asking by de Villiers,[39] quantifiers by Mattei and Roeper,[40] Philip,[41] Crain and Thornton,[42] articles by Schafer and de Villiers,[43] and lexical contrasts by Waxman and Hatch,[44] for example, all suggested still other formats in the different language domains. During the initial National Institutes of Health Screener grant and the first phase of a contract to Seymour for a comprehensive language test, items were adapted from experiments and tried out with AAE learners in Hartford, CT, and Springfield, MA, to establish their utility before being submitted to nationwide field testing.

The goal of the piloting and then the first round of field testing was to narrow down the candidate items to the most effective ones. We had two objectives in mind, Dialect Identification and Diagnosis of Language Impairment; thus, useful items were considered to be those that could be embodied in clear-cut paper-and-pencil-based format and would meet these criteria:

1. they showed steady development across age

For identifier items,

2. they showed a clear distinction between dialect groups

For diagnostic items,

3. they were neutral with respect to the dialect spoken, and
4. there was a clear distinction between typical and disordered performance.

Step 2: Field Testing/Developmental Milestones

The field testing research for this experimental test also served as the data collection for determining developmental milestones for AAE learners. It aimed to establish the milestones of development, a metaphor for the road markers that tell the distances between what are here "landmarks" of language behavior. At what age do we expect a typically developing child to understand how to give an exhaustive response to a double *wh*-question? At what age do we expect most children to use the language for thought to tell about a character's actions and motivations in a short narrative? In addition, by extension, at what age is it a sign of delay not to do so?

Despite progress understanding the unique structures of AAE adult language, no comprehensive picture of AAE child language and the stages of its development has yet emerged.[45–47] Therefore, the developmental milestone research was designed primarily to find the course of mastery in the AAE population for the noncontrastive items in the set; that is, features shared with MAE, such as past tense copula *was/were* or rules for embedding *wh*-clauses. Neither has the prevalence of contrastive items in the speech of AAE-learning children[48] yet been established across a range of ages. Thus, a portion of the developmental milestone research was devoted to determining which AAE features are most persistent, and at what ages and in what contexts AAE features are a significant factor in AAE-child speech.

In theory, children speaking either AAE or MAE should find items based on noncontrastive features equally easy or hard; but there are several reasons why the dialects could differ even on noncontrastive features <u>in development</u>. Features might be similar yet not

identical, or they might not occur with equal frequency. They may interact differently with other features that are contrastive. For example, the tendency for plural nouns to appear without an "*-s*" marking (e.g., "two cup") in AAE may affect the child's likelihood of producing a bare noun as a singular (e.g., "cup" instead of "the cup"). Or, if the third-person /*-s*/ is absent in AAE, then number agreement may be a much less salient property overall. What is the impact of that on the use of quantifiers, such as *every,* which takes a singular noun, versus *all*, which generally modifies a plural noun? The AAE-speaking child may find it easier than the MAE child to treat the two quantifiers alike, and then take longer to recognize the special properties of *every*. It is not clear, or even likely, that equivalence of structures in the adult dialects would always translate into a similar schedule for learning those structures in the child language of both dialects.

Tryout Research/Developmental Milestones Procedures

For this research phase it was important to obtain a diverse sample of 1257 children from across the United States (see Table 1). To accomplish this, 477 speech-language pathologists [footnote to TPC] were recruited to assist in "trying out" more than 300 test items. The items were divided into 14 subdomains, and they assessed more than 30 different language constructs.

Characteristics of the Sample

Given that the original goal of the project was to find a means to identify AAE-speaking children at risk for language impairment as they entered school, the focus of the tryout research was children ages 4 to 6 years, who comprised approximately 65% of the sample. The items were also tested on children ages 7 through 12 years to ensure that ceilings were reached for the different target behaviors. It was also important to test proportional numbers of children of each gender living in different regions of the country to ensure that any patterns found were not limited to only one gender or region.

DISCRIMINATION

Equally important was to discover which items were capable of discriminating disorder through the age range, so the same items were tried out on 250 African American children diagnosed with and receiving services for language impairment. (There was also a group of 147, including 16 additional children not in the main study, who had been identified as phonologically impaired to test the noncontrastive, diagnostic phonology items.) This aspect of the developmental milestones data collection allowed the selection of items that followed a different path of development in the typically developing (TD) and LI children. In the end, a set of items was found that showed steady development through age 9 years (ceiling values were reached at 10 years for the majority of them, even among LI children).

Table 1 Research Sample for the Developmental Milestones Research

	Age (years)									
Characteristic	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
AAE										
TD	108	131	143	15	33	27	37	19	27	540
LI	27	45	50	16	29	15	30	14	16	242
MAE										
TD	60	61	73	16	20	20	21	18	31	320
LI	21	29	34	9	19	12	14	6	11	155
Total	216	266	300	56	101	74	102	57	85	1257

AAE, African American English; TD, typically developing; LI, language impaired; MAE, Mainstream American English.

DIALECT

Because the prevalence of AAE speakers and the density of their dialect was thought to be greatest at lower economic and educational levels,[49,50] the greatest portion of the children (79%) were from families where the parents had only a high school education or less. Finally, given that the items included many innovative elements never before included on a language test, they were also tried out on a comparison set of 475 MAE-speaking children, both TD and LI. These children were matched as closely as possible to the AAE sample with respect to economic and educational level. As in the AAE group, about one third of the children were diagnosed with and receiving services for language impairment. Items that showed different performance between the two dialect groups were eliminated from the pool of candidate items.

There were 49% females overall, more or less evenly distributed throughout the age ranges. (One exception was that the LI children were 61% male; to keep the general balance, the TD groups had more females than males, 55% and 45%, respectively.) The regional distribution of participants was generally based on the distribution of African Americans in the different parts of the country according to the Current Population Survey of the Census Bureau of October 2000 (Table 2).[51]

Table 2 Geographic Distribution of the Research Sample

Sample	North Central	South	Northeast	West
% of subjects	26%	58%	7%	9%
Census figures	25%	51%	15%	9%

RESULTS

The broad results of the milestone research are stunning. They show that typically developing AAE and MAE speakers perform similarly with respect to many sophisticated syntactic and semantic aspects of their grammar.

In analyses of variance, for example, *wh*-questions (WH), articles (AR), and fast mapping (FM), graphed in Figures 1–3, showed extremely large age effects, but no effect of dialect or interaction of age and dialect. [WH: Age, $F(5, 1002) = 46.427$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.380$, $p = 0.538$; Age by Dialect, $F(5, 1002) = 0.385$, $p = 0.859$. AR: Age, $F(5, 1002) = 41.268$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.006$, $p = 0.937$; Age by Dialect, $F(5, 1002) = 0.441$, $p = 0.820$. FM: Age, $F(5, 1002) = 29.600$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.361$, $p = 0.548$; Age by Dialect, $F(5, 1002) = 0.603$, $p = 0.670$.]

In addition, from the point of view of pragmatics in communicative role-taking (CR) and question asking (QA), both groups show that they are developing an age-appropriate sense of how language is used to give and receive specific information (Figs. 4 and 5). [CR: Age, $F(5, 1002) = 46.901$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.025$, $p = 0.875$; Age by Dialect, $F(5, 1002) = 0.620$, $p = 0.685$.

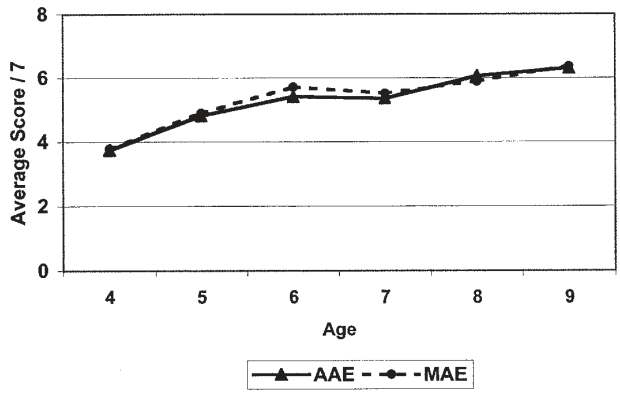


Figure 1 *Wh*-question scores by Dialect group. AAE, African American English; MAE, Mainstream American English.

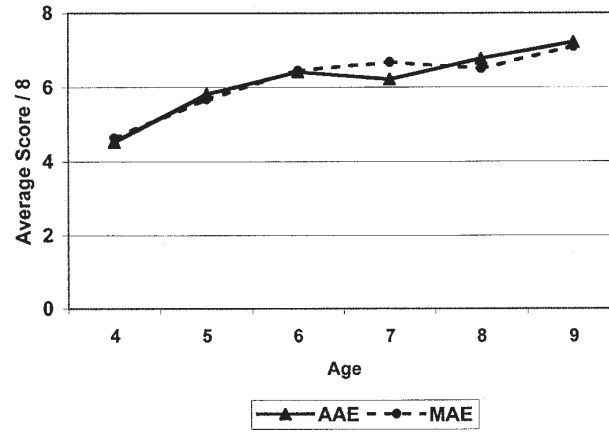


Figure 2 Article scores by Dialect group. AAE, African American English; MAE, Mainstream American English.

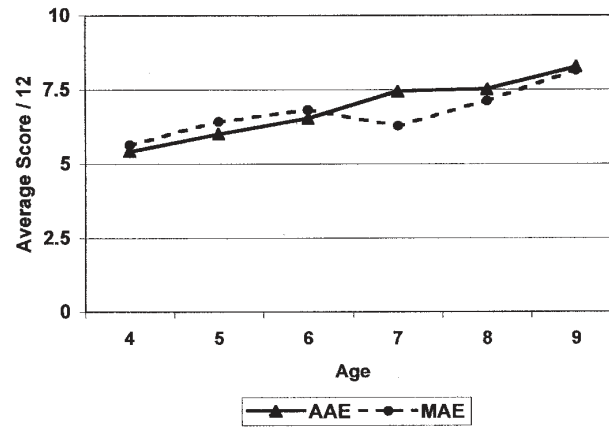


Figure 3 Fast mapping novel verbs by Dialect group. AAE, African American English; MAE, Mainstream American English.

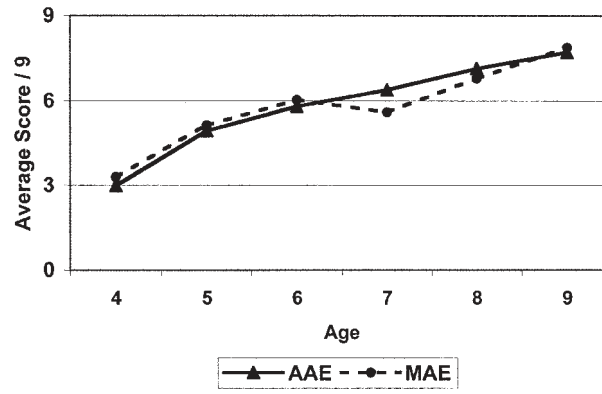


Figure 4 Question asking scores by Dialect group. AAE, African American English; MAE, Mainstream American English.

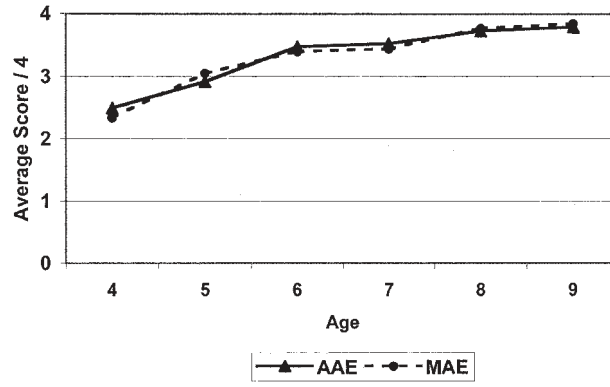


Figure 5 Communicative role scores by Dialect group. AAE, African American English; MAE, Mainstream American English.

QA: Age, $F(5, 1002) = 50.876$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.034$, $p = 0.853$; Age by Dialect, $F(5, 1002) = 0.556$, $p = 0.734$.]

The milestone data also make it clear that there is a small set of predictable morphosyntactic and phonological elements, the language variation score (LVS), that distinguish TD AAE speakers from MAE speakers. In this case (Fig. 6), the strong dialect effect diminishes slightly with age and so one sees a just barely significant interaction. [LVS: Age, $F(5, 696) = 15.748$, $p < 0.0001$; Dialect, $F(1, 707) = 337.615$, $p < 0.0001$; Age by Dialect, $F(5, 696) = 2.274$, $p = 0.046$.] This is not to say that the AAE grammar is characterized only by this small set of features. Rather, these features pattern with the deeper aspects of AAE grammar—negative concord and the rich aspectual and phonotactic systems—so the surface features can be used as proxies for the deeper aspects when it is useful to clarify quickly and efficiently whether the individual is an MAE speaker or not.

Finally, the milestones for these same tasks are significantly different between LI and TD children in all four domains (Figs. 7–10): Syntax (SYN), Pragmatics (PRG), Semantics (SEM), and Phonology (PHO; the comparison for phonology between children with phonological impairment and those with no impairment) (Figs. 7–10). [SYN: Age, $F(5, 1002) = 89.140$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 114.77$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = .765$, $p = 0.575$. PRG: Age,

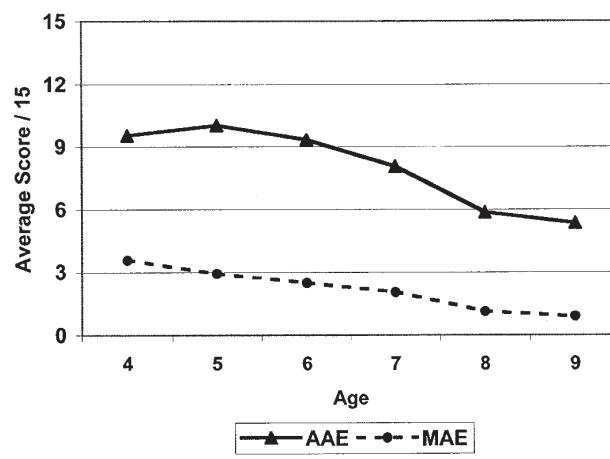


Figure 6 Contrastive elements by Dialect group (typically developing only). AAE, African American English; MAE, Mainstream American English.

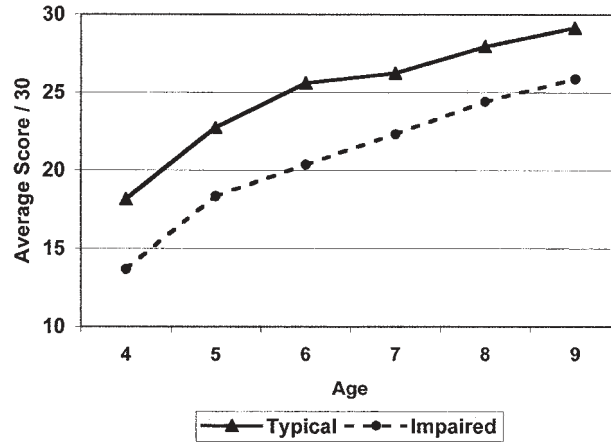


Figure 7 Syntax domain by Clinical Status.

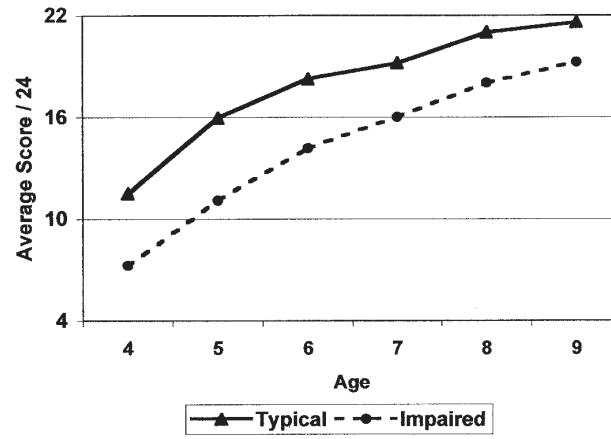


Figure 8 Pragmatics domain by Clinical Status.

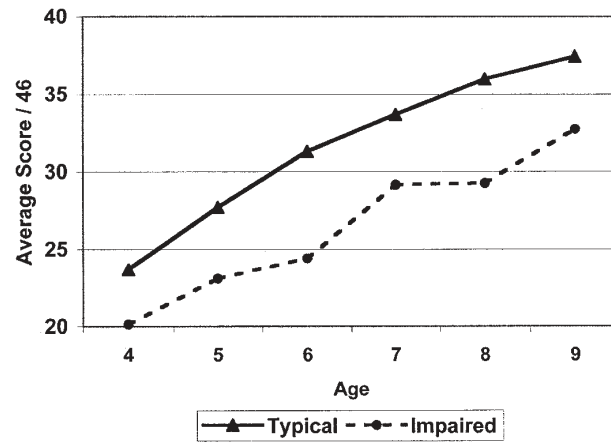


Figure 9 Semantics domain by Clinical Status.

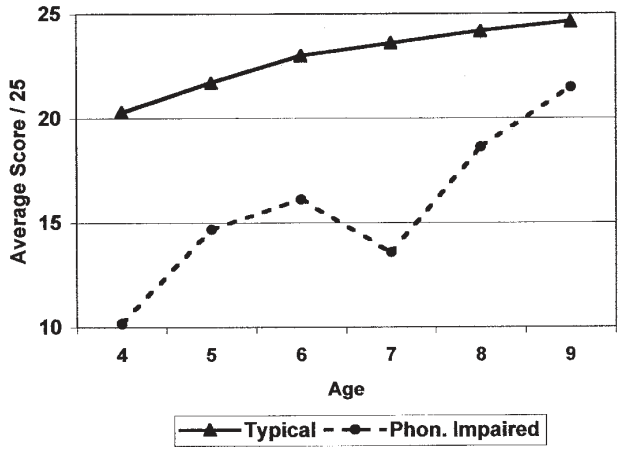


Figure 10 Phonology domain by Clinical Status (articulation disorder). Phon., phonologically (Impaired).

$F(5, 1002) = 108.51$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 110.71$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 1.359$, $p = 0.237$. SEM: Age, $F(5, 1002) = 82.406$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 130.589$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 2.215$, $p = .051$. PHO: Age, $F(5, 1025) = 20.861$, $p < 0.0001$; Phonological Status, $F(1, 1030) = 184.11$, $p < 0.0001$; Age by Phonological Status, $F(5, 1019) = 3.590$, $p = 0.003$.]

MAJOR FINDINGS

The developmental milestones/tryout research described above provided a wealth of data on the language patterns of the four groups of children studied: TD AAE, LI AAE, TD MAE, and LI MAE. The findings on the 4- to 6-year-old children are summarized in the April 2002 report to the National Institutes of Health,[52] portions of which are available from the authors. Additional patterns are being investigated in articles in preparation by Roeper, Strauss, Jackson, Johnson, de Villiers, Pearson, and Velleman, as well as others.[10,11,53]

The major findings from the research we conducted and report on in the articles in this issue can be summarized as follows:

1. A set of 10 contrastive morphosyntactic and 5 contrastive phonological features reliably distinguish AAE from MAE speakers among TD children aged from 4 to 12 years.
2. A set of 17 noncontrastive items (7 morphosyntax, 4 *wh*-comprehension, and 6 nonword repetition) reliably distinguish 4 levels of risk for language delay for children aged 4 to 9 years, regardless of the child's dialect.
3. A set of 25 noncontrastive Phonology items reliably identify risk for speech disorder, regardless of the child's dialect.
4. One hundred sixteen noncontrastive items (46 in Semantics, 28 in Syntax, 17 in Pragmatics, plus the 25 Phonology items) reliably distinguish typical development from language disorder, regardless of the dialect of the child.

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 Assessment, 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. Labov W. The logic of non-standard English. In: Williams F, ed. Language and Poverty. Chicago: Markham; 1970:153–189
2. Wolfram W, Fasold RW. The Study of Social Dialects in American English. Englewood Cliffs, NJ: Prentice Hall; 1974
3. Craig HK, Washington JA. The complex syntax skills of poor, urban, African American preschoolers at school entry. Lang Speech Hear Serv Sch 1994;25:181–190
4. Washington JA, Craig H. Morphosyntactic forms of African American English used by young children and their caregivers. caregivers. Appl Psycholinguist 2002;23:209–231
5. Wolfram W. Dialects and American English. Englewood Cliffs, NJ: Prentice Hall; 1991
6. Chomsky N. Conditions on transformations. In: Anderson S, Kiparsky P, ed. A Festschrift for Morris Halle. New York: Holt, Rinehart, and Winston; 1973:232–286
7. Chomsky N. The Minimalist Program. Cambridge, MA: MIT Press; 1995
8. Prince A, Smolensky P. Optimality Theory: Constraint Interaction in Generative Grammar. Piscataway, NJ: Center for Cognitive Science Rutgers University; Technical Report 2, 1999
9. Archangeli D, Langedoen T, eds. Optimality Theory: An Overview. Malden, MA: Blackwell; 1997
10. Bryant T, Abdulkarim L, Velleman S, Seymour H. Predictors of consonant cluster reduction in African American English. Paper presented at: the Annual Meeting of the American Speech Language Association, Atlanta; Nov. 21–24, 2002
11. Charko T, Velleman S. The influence of dialect on children's phonotactic constraint rankings. Poster presented at: the Child Phonology Conference, University of British Columbia; July 3–9, 2003
12. Roeper T. Diagnosing language variations: underlying principles for syntactic assessment. Semin Speech Lang 2004;25:xxx–xxx
13. Ross JR. Constraints on Variables in Syntax. Dissertation, MIT, Cambridge, MA; 1967
14. Chomsky N. On wh-movement. In: Culicover PW, Wasow T, Akmajian A, eds. Formal Syntax. New York: Academic Press; 1977:71–132
15. Chomsky N. Barriers. Cambridge, MA: MIT Press; 1986
16. Otsu Y. Universal Grammar and Syntactic Development in Children: Toward a Theory of Syntactic Development. Dissertation, MIT, Cambridge, MA; 1981
17. de Villiers JG, Roeper T, Vainikka A. The acquisition of long-distance rules. In: Frazier L, de Villiers J, eds. Language Processing and Language Acquisition. Boston: Kluwer Academic; 1990:257–297
18. Bland L, de Villiers J, Roeper T, Champion T, Seymour HN. Wh-movement in children of divergent language backgrounds. Presented at: the Annual Meeting of the American Speech-Language-Hearing Association; San Antonio, TX; Nov 19–23, 1992
19. Roeper T, de Villiers JG. Lexical links in the wh-chain. In: Lust B, Hermon G, Kornfilt J, eds. Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives. Vol. II: Binding, Dependencies, and Learnability Hillsdale, NJ: Lawrence Erlbaum; 1994:357–390
20. Coles D. Barrier Constraints on Negative Concord in African American English. Dissertation, University of Massachusetts, Amherst, MA; 1998
21. Green L. Topics in African American English: The Verb System. Dissertation, University of Massachusetts, Amherst, MA; 1993
22. Green L. African American English: A Linguistic Introduction. Cambridge, MA: Cambridge University Press; 2002
23. Terry JM. On the Articulation of Aspectual Meaning in African-American English [dissertation] 2002
24. Jackson J. Linguistic Aspect in African-American English-Speaking Children: An Investigation of Aspectual "Be". Dissertation, University of Massachusetts, Amherst, MA; 1998
25. Wyatt TA. Linguistic Constraints on Copula Production in Black English Child Speech. Dissertation, University of Massachusetts, Amherst, MA; 1991

26. Dayton E. Grammatical Categories of the Verb in African American Vernacular English [dissertation]; 1996
27. Leonard L, Eyer J, Bedore L, Grela B. Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. J Speech Lang Hear Res 1997; 40:741–753
28. Rice M, Wexler K. Toward tense as a clinical marker of specific language impairment in English-speaking children. J Speech Lang Hear Res 1996; 39:239–257
29. Leonard LB. Specific language impairment: characterizing the deficit. In: Levy Y, Schaeffer J, eds. Language Competence Across Populations: Toward a Definition of SLI Mahwah, NJ: Lawrence Erlbaum; 2002:209–232
30. Bedore L, Leonard L. Specific language impairment and grammatical morphology: a discriminant function analysis. J Speech Lang Hear Res 1998;41: 1185–1192
31. Leonard L, Miller C, Gerber E. Grammatical morphology and the lexicon in children with specific language impairment. J Speech Lang Hear Res 1999;42:678–689
32. Jakubowicz C. About the nature of the linguistic deficit in SLI. Paper presented at: Symposium on the Acquisition of Functional Categories in French across Learner Contexts, IASCL, University of Wisconsin, Madison, WI; July 16–21, 2002
33. Clahsen H, Hansen D. The grammatical agreement deficit in specific language impairment: evidence from therapy experiments. In: Gopnik M, ed. The Biological Basis of Language. Oxford, UK: Oxford University Press; 1997:148–168
34. Jakubowicz C, Nash L, Rigant G, et al. Determiners and clitic pronouns in French-speaking children with SLI. Lang Acquis 1998;7:113–160
35. Paradis J, Crago M, Genesee F. Object clitics in bilingual and monolingual French-speaking children with SLI. Paper presented at: the Symposium on Morphosyntactic Deficits in Children with SLI: Crosslinguistic and Bilingual Perspectives, IASCL-SRCLD [International Association for the Study of Child Language—Symposium on Research on Child Language Disorders], University of Wisconsin-Madison; July 16–21, 2002
36. Jakubowicz C. What specific language impairment in French (Romance) can tell us about the computational component of the language faculty and its development. Abstract available at http://adur.let.rug.nl/~goins/abstracts/jakubowicz.html
37. Vainikka A, Roeper T. Abstract operators in early acquisition. Linguist Rev 1995;12:275–310
38. Roeper T. The acquisition of implicit arguments and the distinction between theory, process, and mechanism. In: MacWhinney B, ed. Mechanisms of Language Acquisition. Hillsdale, NJ: Lawrence Erlbaum; 1987:309–343
39. de Villiers P. Assessing English syntax in hearing-impaired children: eliciting production in pragmatically-motivated situations. In: Kretschmer R, Kretschmer L, eds. Communication Assessment of Hearing-Impaired Students. Academy of Rehabilitative Audiology, Monograph Supplement Vol 2;1988:41–72
40. Mattei E, Roeper T. On the acquisition of "all" and "some". Papers and Reports on Child Lang Dev 1975;9:63–74
41. Philip W. Event quantification in the acquisition of universal quantification. Dissertation, University of Massachusetts, Amherst, MA; 1995
42. Crain S, Thornton R. Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics. Cambridge, MA: The MIT Press; 1998
43. Schafer RJ, de Villiers J. Imagining articles: What 'a' and 'the' can tell us about the emergence of DP. In: Howell SC, Fish SA, Keith-Lucas T, eds. BUCLD 24: Proceedings of the 24th Annual Boston University Conference on Language Development, Vol. 2. Boston, MA: Cascadilla Press; 2000:609–620
44. Waxman SR, Hatch T. Beyond the basics: preschool children label objects flexibly at multiple hierarchical levels. J Child Lang 1992;19:153–166
45. Seymour HN. The invisible children: a reply to Lahey's perspective. J Speech Hear Res 1992; 35: 640–641
46. Seymour HN, Roeper T. Grammatical acquisition of African American English. In: Taylor O, Leonard LB, eds. Language Acquisition across North America: Cross-Cultural and Cross-Linguistic Perspectives. San Diego: Singular; 1999:109–153
47. Craig HK. The challenges of conducting language research with African American children. In: Kamhi A, Pollock K, Harris J, eds. Communication Development and Disorders in African American Children: Research, Assessment, and Intervention. Baltimore: Paul H. Brookes; 1996: 1–18
48. Craig HK, Thompson CA, Washington JA, Potter SL. Phonological features of child African American English. J Speech Lang Hear Res 2003;46: 623–635
49. Washington JA, Craig HK. Socioeconomic status and gender influences on children's dialectal variations. J Speech Lang Hear Res 1998;41:618–626
50. Stockman IJ. Semantic development of African American children. In: Taylor OL, Leonard LB, eds. Language Acquisition across North America:

Cross-Cultural and Cross-Linguistic Perspectives. San Diego: Singular; 1999:61–109
51. Current Population Survey: School Enrollment Supplemental File [CD ROM]. Washington DC: US Bureau of the Census; 2000
52. University of Massachusetts Working Group on African American English. Development Milestones for AAE Speaking 4-, 5-, 6-Year-Olds. Unpublished report. Amherst, MA; 2002
53. Strauss U, Roeper T, Pearson BZ, Seymour H. Acquisition of universality in wh-questions. Paper presented at: the Generative Linguistics in the Old World Conference. Lund, Sweden; April 9–11, 2003

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. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 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.

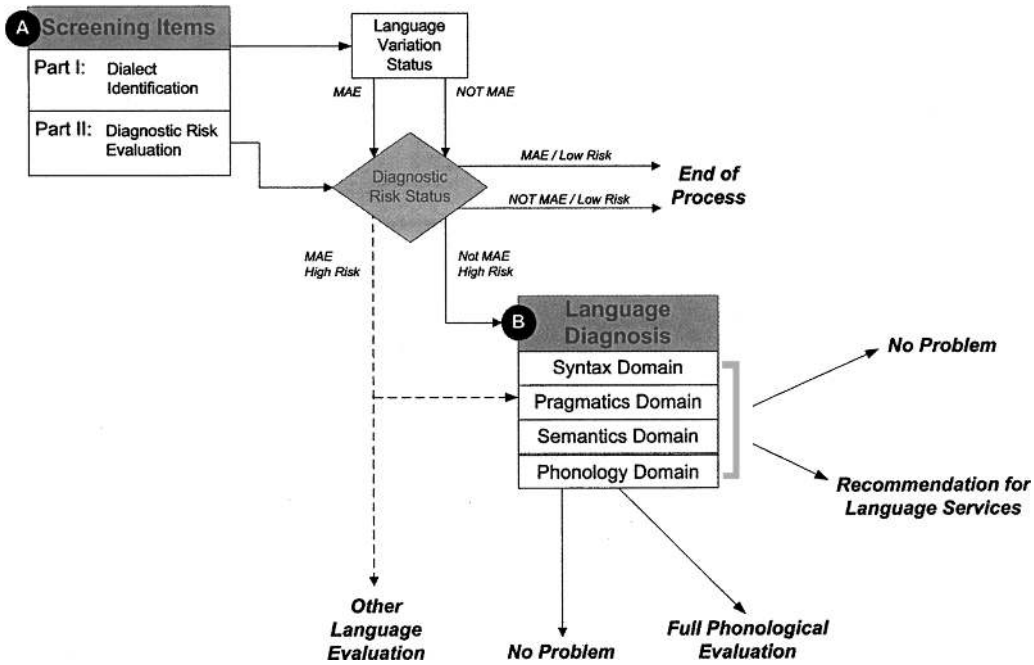


Figure 1 Relationship between elements of proposed assessments. (A) and (B) represent potential starting points for the assessment. MAE, Mainstream American English.

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 non-contrastive 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)

Dialect Identification versus Evaluation of Risk in Language Screening

Lois Ciolli, M.A., C.C.C.-S.L.P.,[1] and Harry N. Seymour, Ph.D.[2]

ABSTRACT

This article proposes that any dialect-neutral screening test should consist of two parts: one part using contrastive items to screen for Language Variation Status (Mainstream American English [MAE] or a degree of variation from MAE), and a second part using noncontrastive items to screen for degree of risk for language disorder (low, medium, or high). The two scores are interpreted together in forming an overall clinical profile of a child. The implications for further diagnostic testing of both MAE and non-MAE speakers who exhibit risk for language disorder are discussed. Most importantly, for the African American child, viewing the results of the two types of proposed screening items together clears the "variation due to speech and language dialect" issue out of the way so that if further evaluation is needed, it is done for valid reasons, not superficial ones.

KEYWORDS: Language Variation Status, Diagnostic Risk Status, degree of risk, variation from mainstream, optimal, obligatory, contrastive aspects of language, noncontrastive aspects of language, risk for disorder

Learning Outcomes: As a result of this activity, the participant will be able to (1) describe contrastive aspects of language that can be used to identify African American English (AAE) dialect usage in children ages 4 to 12 years, and (2) describe noncontrastive aspects of language that can be used to determine degree of risk for a language disorder in children ages 4 to 9 years, regardless of whether AAE or Mainstream American English (MAE) is spoken.

DIALECT IDENTIFICATION VERSUS RISK DIAGNOSIS

If the question for a screening procedure is to identify dialect versus disorder in African American English (AAE) speakers, both goals can be accomplished with phonology and morphosyntax items. For example, in phonology, consonant clusters when in word-final position

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: Lois Ciolli, M.A., C.C.C.-S.L.P., The Psychological Corporation, Harcourt Assessment, Inc., 19500 Bulverde Road, San Antonio TX 78259. E-mail: lois_ciolli@harcourt.com. [1]Senior Research Director, The Psychological Corporation, Harcourt Assessment, Inc., San Antonio, Texas; [2]Professor Emeritus, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01, 033,040,ftx,en;ssl00181x.

have different rules in AAE and MAE. Therefore, if the purpose is to establish Language Variation Status, the phonology items should target only the final position, where differences between AAE and MAE are greatest. In contrast, in screening for phonological disorder, it is reasonable to present targets only in initial and medial positions, where AAE and MAE speakers (for the most part) treat them alike.

Similarly, in morphosyntax, the distinct sets of dialect identifier and language disorder screening items can both include items involving the past tense verb *was* (or *were*). For the identifier (Language Variation Status) objective, the focus of these items should be the nature of the subject-verb agreement, which differs reliably across the dialects: in AAE, *was* occurs predictably with plural *they* ("they was"), whereas MAE speakers almost exclusively say "they were." In contrast, the items in a screener for language disorder should focus on whether *was* is deleted or not; neither dialect would permit deletion in that context.[1,2] In the present tense, AAE forms of *to be* are used optionally,[3] but past tense *was* or *were* are obligatory in AAE because they carry the tense marking, information essential to understanding the sentence. If a child of either dialect deletes *was*, that would be a marker of disorder. Thus, the contrastive agreement pattern for *was* can be used in determining Language Variation Status, but the noncontrastive obligatory presence of *was* can be used for determining risk status.

Item Goals

PART I: DEVELOPING A LANGUAGE VARIATION SCREENER

We constructed a potential set of items with the goal to *maximize* the difference in responses between the AAE and MAE speakers. Given that the most obvious contrastive features between the two dialects are morphological and phonological, those features were the focus of research for this part of the battery we developed. Out of 182 morphosyntax and phonology items that were tested during the preliminary research, 15 items that best differentiated the two dialect groups were selected for the final version of our proposed screening test. Although there are many other features typically considered characteristic of AAE, they are not represented in our final version (e.g., multiple negation, "he don't have no shoes"). The items selected were those that best differentiated the two dialect groups in morphological and phonological patterns at all of the ages we looked at. The end result was 5 phonology and 10 morphosyntax items that seemed quick to administer and score.

We found that the most discriminating phonology items involved only three phonemes—voiceless "θ," voiced "ð," and the consonant cluster /ft/. These were incorporated into five phonology items as stimulus targets and were designed to elicit AAE patterns in the final position of words: /θ/ → [f]; /ð/ → [v]; /ð/ → [d]; /ð/ → zero; and /ft/ → [f]. In administering these five phonological targets, the examiner says a short sentence about the picture and asks the child to repeat it. For example, the examiner might show the child a picture of a child's face with her mouth open and say, "I see she can open her mouth."

The next 10 items that we field tested focus on morphosyntactic aspects of the child's speech. For these items, the child is asked to look at a picture and finish a sentence or answer a question about it. The items assess a child's marking of third-person, present tense, singular verbs (*has, -s, -es, does*) and the third-person, past tense, plural verb agreement, as in "they was" or "they were."

A child's dialect patterns can be easily identified with this set of proposed items. We have found that a child who is a speaker of AAE will respond to most of the 15 target items in a manner typical of AAE. Likewise, in response to the same 15 items, a child who is a speaker of MAE will produce more MAE patterns. Nonresponses or response patterns not typical of either the predicted AAE or MAE targets also could be recorded, but not entered into the scoring. The combined score for the phonology and morphology items can be used to determine a child's variation status as either "MAE," "Some Variation from MAE," or "Strong Variation from MAE." As seen in

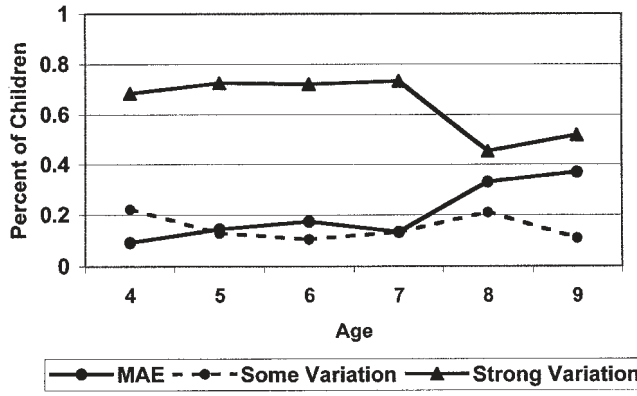


Figure 1 Language variation categories (African American English [AAE]-background children). MAE, Mainstream American English.

Figures 1 and 2, our experimental work shows that the dialect groups have very different numbers of children in each of these categories.

Typically developing children from MAE family backgrounds rarely show any Language Variation Status except MAE (Fig. 2), whereas children from AAE family backgrounds show all three patterns (Fig. 1). Even at the older ages when many of the children are learning to code-switch into MAE, typically developing African American children are still more likely to speak with some or strong variation from MAE.

Item Goals—Part II: Diagnostic Risk Status

We next sought items that were effective in identifying children at risk for a language disorder and in eliciting the same responses from typically developing children, regardless of whether they were AAE or MAE speakers. Out of the 313 items that were researched in the second phase of piloting, we found 17 items that as a group best discriminated between the typically developing children and those with language impairments in the research sample.

Five of the items focus on the child's production of the third-person singular past tense verb *was*, in either copula or auxiliary form. Two more of the items that we identified for this function also focus on morphosyntax; these are possessive pronouns. Four more items involve *wh*-questions and are discussed by Roeper.[4] Again, both the typically developing AAE and MAE speakers understood these in the same manner.

The last set of items that we developed arose from work done by Campbell et al[5] on nonword repetition abilities of children with and without specific language impairment

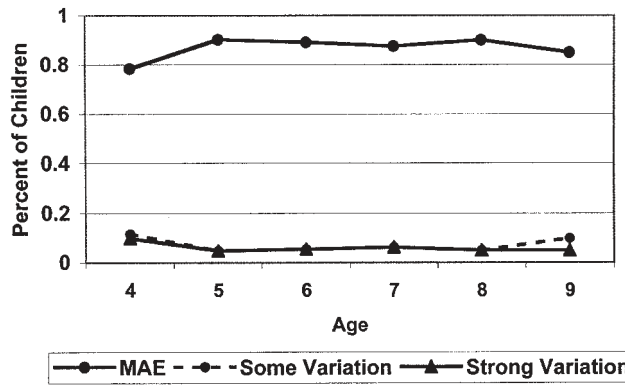


Figure 2 Language variation categories (Mainstream American English [MAE]-background children).

(SLI). Nonword repetition tasks have shown to be useful in the identification of SLI children without biasing effects toward speakers of AAE. It is believed that this task's effectiveness derives from the processing-dependent nature of the task, as opposed to the usual language assessments that are heavily language dependent. We chose six targets for this task, in which the child is asked to repeat a nonsense word two to four syllables long, such as the trial item *poo-zle*. All six of the actual items were carefully constructed to account for the phonotactics of AAE to ensure that AAE speakers would not be placed at a disadvantage when attempting to repeat them. The nonwords were constructed out of single phonemes (no consonant clusters) and used only those final consonants that in our archive of AAE child speech were least likely to be omitted by AAE speakers.[6,7]

We believe that the child's risk status can be easily determined from the response patterns on these 17 diagnostic items. Based on a combined score for all items, the child could be classified as at "the lowest risk for language disorder," "low to medium risk," "medium to high risk," or "the highest risk for disorder." Because this is done in a dialect-neutral way at the outset, it sets the framework for further testing. Typically, in an effort to accommodate divergent dialects such as AAE, speech-language professionals (SLPs) modify standardization guidelines of language tests by adjusting scores or interpreting results in accordance with the child's dialect. Such practices are highly questionable because they violate the tests' standardization protocols. By contrast, our proposed screening items do not differ for AAE and MAE speakers, and scoring guidelines can be the same for both groups.

What we have chosen to call the Diagnostic Risk Status is indicated by a score that reflects the number of elements indicative of risk. Some responses, like omitting the past tense *was* are weighted double because they are particular warning signs or "red flags." The higher the score, the greater the risk. The difference between the typically developing and impaired groups in both dialect samples is shown in Figures 3 and 4.

The effects of development and clinical status is significant for all children, although the gap narrows with age, as the significant interaction indicates: Age, $F(5, 1002) = 56.465$, $p < .0001$; Clinical Status, $F(1, 1002) = 168.07$, $p < .0001$; Age by Clinical Status, $F(5, 1002) = 3.172$, $p = .008$. The gap appears even greater between the clinical groups in the MAE background children: Age, $F(5, 362) = 23.927$, $p < .0001$; Clinical Status, $F(1, 362) = 114.95$, $p < .0001$; Age by Clinical Status, $F(5, 362) = 2.316$, $p = .046$.

INTERPRETING PERFORMANCES ON OUR SCREENING ITEMS

The two scores on our screening items for language variation and diagnostic risk can be interpreted together in forming an overall clinical profile of a child. It is possible to obtain four major performance profiles: MAE/lowest risk; MAE/some risk; some variation from

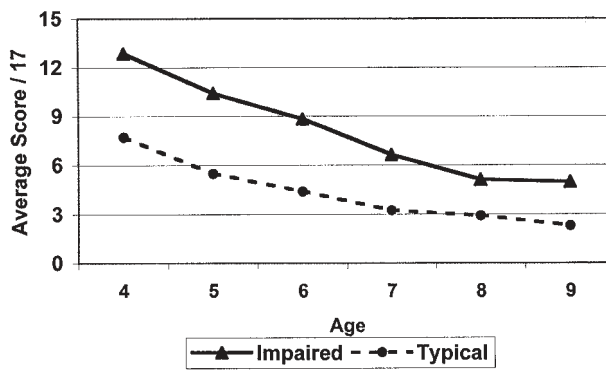


Figure 3 Noncontrastive diagnostic elements by clinical status.

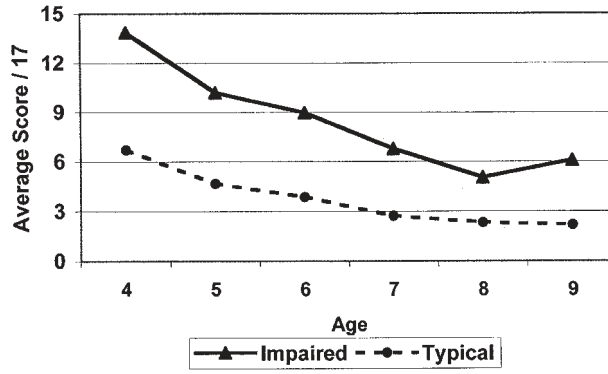


Figure 4 Noncontrastive diagnostic elements by clinical status (Mainstream American English [MAE]-background children only).

MAE/lowest risk; some variation from MAE/some risk. Each of these profiles is discussed below with interpretative implications.

MAE/Lowest Risk

The child who falls into this profile has produced all or most of the 15 variation items and all or most of the 17 diagnostic items in a manner consistent with MAE. We believe no further clinical action should be necessary.

MAE/Some Risk

The child who fits this description is a candidate for further diagnosis to confirm that a language disorder exists and to determine the nature of the problem. He or she has responded to the 15 variation items in the MAE manner, but has done so on relatively few of the 17 diagnostic items. The child's performance on the 15 variation items suggests few difficulties with morphosyntax. Therefore, further diagnosis could focus on other aspects of language along the lines discussed in the articles in this issue (especially Roeper,[4] P. de Villiers,[8] J. de Villiers,[9] and Seymour[10]).

Some Variation from MAE/Lowest Risk

Children with a certain number of predictable AAE responses on the 15 variation items scored in either the Some or Strong Variation from Mainstream American English category. These children were considered speakers of AAE. Because many of the AAE responses can be heard in the speech of young children, regardless of dialect, it is possible for non-AAE-background children to use some of the very patterns that indicate AAE status (e.g., absent third-person *-s*, as in "he talk"). However, we have found that younger children produced enough responses indicative of an AAE pattern that the distinctions between AAE and MAE typical development were observed even at the younger ages.

The important fact about this profile is that AAE children have been identified and not penalized for their use of AAE features. Their performance on the 15 variation items confirmed their status as AAE speakers, but their predicted MAE responses on the 17 diagnostic items showed the lowest risk for a disorder. We believe it is this Some Variation from MAE/Lowest Risk profile that is most often confused for disorder in other language tests and that an adequate screener must be able to distinguish it from the other the profiles.

Some Variation from MAE/Some Risk

Of all the profiles this is the most complicated to interpret. The interpretation is clear with respect to the clinical purpose to identify impaired children and not penalize typically developing African American children for speaking AAE. Our screening items appear to achieve this most important objective. The complexity arises from the potential overlap in variation status between AAE children at risk

and MAE children at risk. The variation status is relatively straightforward when African American children show AAE patterns on the screener and pass the diagnostic items (i.e., they are typically developing AAE speakers), and when typically developing MAE speakers confirm their variation status by their performance on the variation items. However, what is not clear is when children of either dialect show responses consistent with AAE patterns on the variation items of the screener and also fail the diagnostic items. Because an impaired language system is so unstable, it may show patterns similar to the 15 variation items (deletions such as third-person /-*s*/, *do/does*, and *have/has* substitutions and certainly immature phonological patterns such as [f] substituted for /θ/) Consequently, both AAE and MAE language-disordered children may appear to be AAE speakers. In fact, in the field testing sample, more than half of the children from MAE-speaking backgrounds who were identified for and receiving language services spoke with some or strong variation from MAE.

For the child from an MAE background, these problems on the variation items (contrastive features) could be an indication that something is amiss linguistically. But the stronger evidence *in either dialect* for a language problem will be the child's performance on the diagnostic items (noncontrastive features). As Figure 3 shows, the noncontrastive features distinguish the typically developing from language-impaired children, but they are an even better diagnostic for MAE background children than for children who speak with strong difference from MAE (Figure 4). This is consistent with the observation made by Seymour[11] in the introduction to this issue that normal AAE features can appear similar to patterns of language disorder in MAE speakers. The linguistic pattern of an African American child (or a child from another racial or ethnic background) who has a language disorder and comes from a home and community where MAE is the dominant dialectal pattern will mimic those of an AAE speaker. When the features reflect *accurate* learning of the language model provided to the child, they are markers of dialect. When they are an *inaccurate* reflection of the child's language model, as in the case of a language-impaired child from an MAE-speaking home, then they generally signal disorder.

Thus, the AAE identifiers work best in preventing children who use AAE patterns and who are typically developing from being misdiagnosed. However, these AAE identifiers, when used as the sole source for dialect identification, can be ambiguous with respect to dialect for children whose system is impaired. Our work attempts to overcome this ambiguity by not relying on precisely those contrastive features for overall assessment. Therefore, when AAE status is suspected for an impaired child based on the variation items, other confirming evidence is necessary, such as family and community background. Some evidence of AAE or MAE features may be seen in the child's longer responses to other items, particularly in the Pragmatics domain of the proposed comprehensive language test (P. de Villiers[8]). Indeed, SLPs know that a single test score is only one piece of the assessment puzzle: To figure out what is going on with a child, it is essential to consider all pieces of information before drawing conclusions that affect a child's future educational plans.

FURTHER DIAGNOSTIC TESTING

Because our proposed screening items do not constitute a full diagnostic test, children identified as at risk should be given additional testing. For many reasons discussed throughout this issue, the proposed dialect-sensitive language test described in this issue is recommended. But no test replaces the need for ongoing diagnosis. This is especially important for certain aspects of language that have not been examined by the proposed test.

Although criterion-referenced items can form a comprehensive test of language comprising syntax, semantics, pragmatics, and phonology, the one we propose does not directly examine morphology and this aspect is examined only in a limited way in our screening items. The reason for avoiding morphosyntax to this degree is to prevent disadvantage to AAE speakers. However, because morphology has a high probability of being aberrant when a child's language is impaired, it should not be ignored in subsequent stages of diagnosis

and treatment. Therefore, the following considerations are pertinent to further diagnosis of morphology.

If the child is an MAE speaker who fails our diagnostic screening items and shows difficulty with the variation items by scoring as Some or Strong Variation from MAE, this variation status could be an indication of problems with morphosyntax. There are several options available for testing morphology with MAE-speaking children who give evidence of problems with morphology, and SLPs should follow traditional best practices.[12–15]

In the case of a non-MAE speaker, the objective of assessing morphology would be to determine whether the child's patterns are (1) consistent with noncontrastive targets and (2) typical of AAE speakers in either structure or frequency. Examples of noncontrastive morphological structures would be possessive pronouns (*his*, *hers*, *ours*) past tense copula (*It was cloudy*), or presentational copula (*It's the President*).[16] AAE speakers should not show difficulty with these items unless they have problems with morphology. To determine whether patterns are typical of AAE, there should be an assessment of the frequency of usage. A defining pattern of AAE morphological inflections is optional use, which means they are absent some of the time and present at other times. For example, in the University of Massachusetts archive of AAE child speech, the *is* copula is typically absent about 20% of the time. The exact linguistic contexts that govern its presence and absence are not completely known, but a child who is deleting the *is* copula most of the time is probably showing difficulty.

Most importantly for the African American child, viewing the results of the two types of proposed screening items together clears the "variation due to speech and language dialect" issue out of the way so that if further evaluation is needed, it is done for valid reasons, not superficial ones.

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. Baugh J. A Re-Examination of the Black English Copula. Austin, TX: Southwest Educational Development Laboratory; 1979
2. Labov W. Contraction, deletion, and inherent variability of the English copula. Language 1969; 45:715–762
3. Wyatt TA. Linguistic Constraints on Copula Production in Black English Child Speech [dissertation]. University of Massachusetts Amherst; 1991
4. Roeper T. Diagnosing language variations: underlying principles for syntactic assessment. Semin Speech Lang 2004;25:41–56
5. 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
6. Abdulkarim L, Bryant T, Seymour HN, Pearson BZ. Consonant clusters in child African

American English phonology. Poster presented at: the Annual Meeting of the American Speech-Language-Hearing Association, San Francisco, November 18–21, 1999
7. Abdulkarim L, Bryant T, Seymour HN, Pearson BZ. Final consonants in child African American English phonology. Poster presented at: the Annual Meeting of the American Speech-Language-Hearing Association, San Francisco, November 18–21, 1999
8. de Villiers PA. Assessing pragmatic skills in elicited production. Semin Speech Lang 2004;25:xxx–xxx
9. de Villiers JG. Cultural and linguistic fairness in the assessment of semantics. Semin Speech Lang 2004;25:xxx–xxx
10. Pearson BZ, Ciolli L. Distinguishing dialect and development from disorder: case studies. Semin Speech Lang 2004;25:xxx–xxx
11. Seymour HN. The challenge of language assessment for African American English-speaking children: a historical perspective. Semin Speech Lang 2004;25:3–12
12. Muma JR. Language Handbook: Concepts, Assessment, Intervention. Englewood Cliffs, NJ: Prentice Hall; 1978
13. Bloom L, Lahey M. Language Development and Language Disorders. New York: John Wiley; 1978
14. Nelson N. Childhood Language Disorders in Context: Infancy through Adolescence. Columbus, OH: Merrill; 1993
15. Paul R. Language Disorders from Infancy through Adolescence: Assessment and Interventions, 2nd ed. St. Louis, MO: Elsevier Science; 2001
16. Benedicto E, Abdulkarim L, Garrett D, Johnson V, Seymour H. Overt copulas in African American English speaking children. In: Greenhill A, Hughes M, Littlefield H, Walsh H, eds. Proceedings of the 22nd Annual Boston University Conference on Language Development. Somerville, MA: Cascadilla Press; 1998:50–57

Diagnosing Language Variations: Underlying Principles for Syntactic Assessment

Thomas Roeper, Ph.D.[1]

ABSTRACT

The assessment of complex aspects of children's syntactic development can be carried out in a dialect-neutral fashion. The item types proposed for this purpose encompass *wh*-questions, passives, and articles, and test the child's understanding of implicit grammatical relations, the rules governing syntactic movement, and discourse linking. We present the rationale for the items and the specific research supporting them, as well as some suggestions for how to help children who fail on the concepts to gain a better understanding of them.

KEYWORDS: Implicit information, discourse linking, double *wh*-questions, *wh*-barriers; definite versus indefinite articles, passive, hidden agents, exhaustive set, singleton answers, medial question

Learning Outcomes: As a result of this activity, the participant will be able (1) to identify three types of items that can be included in a linguistically fair assessment of syntax, and (2) to identify key concepts important to the development of the syntax items on the proposed assessment.

Grasping grammar involves moving to an abstract level that is so natural we do not realize how far it is from common sense. Its naturalness flows from the fact that much of it is innate, like vision. Yet like vision, either big problems (complete blindness) or tiny ones (imperfections of focal length) can be present. The challenge of communication disorders is to identify the deeper factors that lie beyond superficial description. Our presentation does not focus on dialect variation, but rather on those features of English that are constant across dialects. They therefore offer the opportunity for dialect-neutral diagnosis of language disorder. We begin with some simple examples and move to the kinds of sentences that may cause children to face real challenges and may cause real failure.

A linguistically fair syntax assessment can include these item types: *wh*-questions,

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: Thomas W. Roeper, Ph.D. Professor, Department of Linguistics, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: roeper@linguist.mass.edu. [1]Professor, Department of Linguistics, University of Massachusetts Amherst, Amherst, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01,041,056,ftx,en;ssl00182x.

passives, and articles. Our focus falls on three key concepts in these domains:

1. Implicit grammatical relations: How do we grasp silent, elliptical knowledge buried in short sentences with much left out?
2. Core properties of questions: the rules and restrictions for syntactic movement within and across clauses and the requirement to answer exhaustively.
3. Discourse linking: How do we make connections across sentences?

These are among the major concepts that can be evaluated by the syntactic assessments we have been exploring, but they represent only a few of the grammatical concepts that a competent speaker must master. They involve many subtle features of everyday language, which are often the missed signals that make conversations go awry.

Following the section on key concepts, we explain the rationale for the items on our assessment, how they are constructed, and some of the specific research results that support their use. We finish with some suggestions about how to elaborate these concepts to help the children who fail on the items gain a better understanding of them.

KEY CONCEPTS

Implicit Grammatical Relations: Ellipsis

We often are unaware that much of what we think we say actually is not stated. If someone says, "Here are some berries. Do you want some?" and then someone asks what was said, one might reply "I asked if you want some berries." However, that was never actually said, only "Do you want *some*?" The hearer must know to fill in the missing "berries."

Parents' speech is constantly elliptical, even with very young children. Here is a dialogue with a 2-year-old child from the CHILDES database[1]:

Mother: There isn't any tapioca.
Child: Have milk.
Mother: There isn't any.
Mother: We'll make some this afternoon.
Child: (unintelligible) make some (unintelligible).

Both *any* and *some* could refer to an unspoken *milk*, but upon reflection, "make some" seems to refer to *tapioca*. The child (and we) must figure it out—and it looks like the child may be having trouble. These are elementary examples, but the possibility for confusion is clear.

Ellipsis (Implicit Grammatical Relations) in Passive Sentences

The need to reconstruct missing information also arises if I say:

1. The bike was stolen.

If someone asks what was said, one might say, "I just said that someone stole the bike." But actually, *someone* was never mentioned and we had to supply the missing agent of *stolen*. This may seem inevitable in a sentence such as (1), because there cannot be any stealing without someone to do it. But it is not inevitable in a contrast such as:

2. The apple dropped.
3. The apple was dropped.

Only in sentence (3) is a missing agent implied. Unlike *steal*, it is not the verb *drop* that demands an agent, but the passive construction (*was* + *ed*). Moreover, the missing agent cannot be the same as the object. If I say:

4. John was being washed.

The grammar of the sentence implies that a missing agent is present who is someone other than John (even though John could wash himself). This is a property that we call "*disjoint reference*" and it means simply that the subject and object are different.

Properties of Questions: Exhaustivity

A second key concept involves what we have called the "bottomless nature" of question words and other words that involve "quantifiers." If someone asks, "What is in your living room?" we could answer with a single word: "furniture." But we could also enumerate every object, or even describe the air, the windows, the views, the people, or colors. Unless

modified by context, questions in principle demand an answer that is full and exhaustive.

These criteria can be captured in the term *exhaustive set*. Imagine that you are in court and the lawyer asks, "Who was in the car the night of the murder?" If there were three people and you answer with just one, you are liable for perjury. The question requires that we exhaustively mention everyone in the set of people who were in the car.

Children may fail to grasp that a question calls for an exhaustive set answer. A competent speaker, regardless of dialect, needs to understand that the word *who* or *what* by itself carries the set interpretation.

Properties of Questions: Movement

A question achieves a focus by an operation called movement. In the following example, we move the question word *what* to the front:

I see a man, a woman, and a horse. →
What do I see (-)?

We can do the same thing with a complex sentence and move *what* over two clauses:

Complex: What did she say she saw (-)?

Here we have bumped into one of the deepest properties of grammar: potential infinity. We could just keep on going:

What do you think I think the teacher said that
Mary told John she saw (-)?

Here we have expanded the same question over five clauses, and it still refers to the object of *see*. It is precisely the ability to swallow effortlessly this kind of potential infinity that is the essence of language ability.

Movement, Ambiguity, and Barriers to Movement

There can be even greater complexity in these structures. The question word can be potentially infinitely far away from the position it started from, as with *see* above, and it can also be ambiguous. In the following sentence there are two different possible *when* sequences:

When did she think (-)
she lost her purse (-)?

We could be asking when she said it or when she lost it. This can, of course, be an infinitely repeatable ambiguity as well:

When did you say (-) she thought (-) she lost
it (-)?

It may appear as if we are complicating the situation and not the syntax, but it becomes clear that these ambiguities are regulated by syntax when we insert another question word:

When did you say (-) *how* she thought she lost it?

A bit of reflection reveals that now all of the "lower" *when* positions are completely cut off and the question refers only to when the saying was done. One question word cannot jump over another. The second question word is called a "barrier."

Although such concocted sentences may seem strange to the reader, they are the stuff of daily life for children. Parents and children use these embeddings readily. One can easily imagine a 5-year-old child saying a sentence to a sibling that has three embedded clauses in it:

Mom said you said I did it, but I didn't.

Articles and Discourse Linking

Finally, we need an intuitive grasp of what articles do. If I say:

John bought a hat. The hat was green.

The *the* tells us that the bought hat and the green hat are the same. Once we introduce a noun with an *indefinite* article (*a*), we can refer back to it with a *definite* article (*the*). Again, this is far from inevitable. Many languages have no articles and so the connection has to be made differently.

One might think that common sense tells us the connection is there. But, actually, that kind of common sense easily wanders into

significant error. Compare the following two sequences:

I bought some cats. The cats are strange.
I bought some cats. Cats are strange.

Without *the*, we refer to cats in general; with *the*, we refer to exactly the cats we bought, (who might be unusual in being strange).

The same connection can be achieved via our knowledge of the world. We can use a definite article to refer to a part of something already mentioned, even though the part itself has not been mentioned. We can say:

Take a cup. Use the handle.

In this context, we mean "the handle of the cup." If we do not know the object, we can still use *the* to assume a part/whole relation. If I say:

John has a motor. The brinch is broken.

From these sentences, we are led by the article to assume that "the brinch" is a part of the motor. This, again, is the sort of knowledge that a competent speaker must have, but is not taught (see Schafer and deVilliers[2] for a systematic review).

CONSTRUCTING A DIALECT-SENSITIVE SYNTACTIC PROBE

Wh-Questions

The *wh*-question comprehension items of the syntax probe are of three types: double *wh*-questions, embedded clauses (with false complements), and barrier questions. The double *wh*-questions involve exhaustivity, the embedded clause questions involve knowledge of implicit relationships and movement of the question word from the position of the constituent it replaces, and the barrier questions involve constraints on movement.

DOUBLE-*WH*: TWO QUESTIONS, TWO EXHAUSTIVE SETS

It is possible to ask two questions at once. With three words,

Who bought what?

we utter a sentence that calls for two answers. The answers each require a set and seek exhaustivity. In addition, a third property must be honored: *pairing*. The question is answered as a paired list:

He bought fruit and she bought vegetables.

Questions of this kind require reference to *all the members* in the *two sets* in *an ordered relation*:

Person 1 bought Thing 1,
Person 2 bought Thing 2.

Any questions can be used, for instance:

How did she catch what?

Our exploratory test (See Acknowledgments) investigated children's performance on this kind of question with stories of this type that make each pair somehow dramatic or interesting. One example constructed to test this concept might read: "This girl caught different things in different ways. She caught that crab with a net and the fish with her fishing pole. (pause) How did the girl catch what?"

RESULTS: TYPES AND AMOUNTS

We obtained three types of typical answers to such questions

1. Paired, exhaustive responses (correct):
   - "She caught the crab with a net and the fish with a pole."
2. Singletons (incorrect):
   - One element: "a crab" "with her pole"
   - Both direct objects, no instruments: "crab and fish"
   - One pair: "the crab with a net."
3. Other:
   - "She fished a lot." "She was playing."

To see how revealing these questions are, we will provide a general overview of results from the field testing.

The difference between typically developing (TD) and language-impaired (LI) children shown in Figure 1 is statistically significant. In the analyses of variance, the main effects of development (age) and disorder (clinical

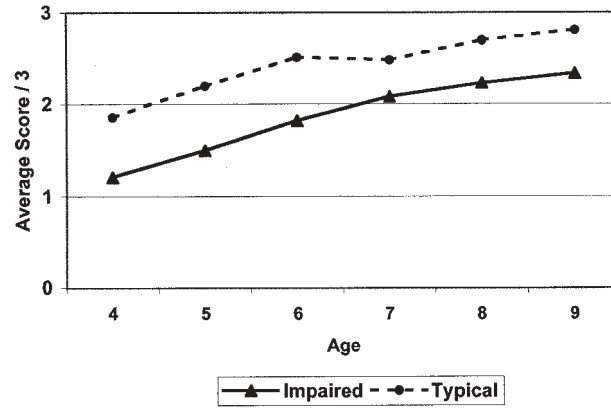


Figure 1 Double *wh*-questions by Clinical Status.

status) are large and consistent [Age, $F(5, 1002) = 18.376$, $p < .0001$; Clinical Status, $F(1, 1002) = 49.966$, $p < .0001$; Age by Clinical Status, $F(5, 1002) = .449$, not significant]. We see that typically developing children can give appropriate answers two thirds of the time at the age of 4; disordered children give an appropriate answer one third of the time and remain consistently behind through the age of 9.

Figure 2 shows that African American English (AAE) and Mainstream American English (MAE) speakers treat these structures essentially the same [Age, $F(5, 1002) = 16.145$, $p < .0001$; Dialect, $F(1, 1002) = 6.435$, $p = .01$; Age by Dialect, $F(5, 1002) = 1.061$, $p = .381$]. Although the difference between the dialects was statistically significant, the simple effect at each age was not reliable except for ages 7 and 9, for which the number of subjects was very small.

Thus, these rare sentences achieve our goal: they produce a method to recognize and identify a disorder without the diagnosis being affected by dialect in a critical age range. In addition, of course, they identify a disorder that has hitherto not been seen: an inability to handle a double question. The inability to handle the double question is pertinent to all questions because it forces out an important feature of questions: the reference to a potentially infinite set. The children who cannot answer double questions will be those who inappropriately give a single response to simple questions in ordinary life, such as the child who mentions just one ingredient when you ask, "What do you need to make a cake?"

CASE STUDIES

Here are two example children from our field testing who give different responses to double *wh*-questions:

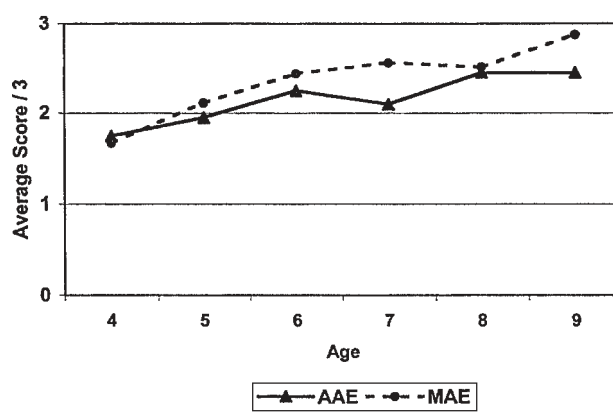


Figure 2 Double *wh*-questions by Dialect.

Child A	Child B
banana and a apple	The Mom ate the apple and the Dad ate the 'nana
fish and crab	fish with her pole, crab with a net

Only child B sees the requirement of a paired answer.

HIDDEN FALSE-CLAUSE QUESTIONS

Complex sentences make substantial cognitive and semantic demands, and require the child to understand implicit relationships across clauses. When one clause is inside another clause, the notion of the truth of the sentence must refer to the whole, not the parts. Thus, if I say,

Who did John say stole the bread?

I may not have said who actually stole the bread. If a child were to treat such a sentence as a conjunction, then both parts would require *true* answers—we would really be asking who did it, not just who John said did it:

Who did John say and (who) stole the bread?

The child must see that putting one verb inside the other changes the truth value: Now we do not have to know or say who stole the bread, just who John said did it. Children definitely can stumble over this computation.

When we answer an embedded question like this, our answer must address the verbs from both clauses: *who John say-steal*. If our answer is based on just one of the verbs (e.g., *who stole?*), we show that we have not understood the relationship of the embedding.

Here is how we presented this kind of question in our field testing:

This girl snuck out one night when her little sister was asleep and brought a pretty plant as a surprise. The next day the little sister saw the bag from the store and asked, "What did you bring me?" The older girl wanted to keep the surprise until later so she said, "Just some soap." (pause) What did the girl say she brought?

Here are typical answers to false-clause questions:

1. "Long-distance" two-clause responses (correct):
   - She said she brought soap (*she say-buy*)
2. One-clause responses (incorrect):
   - (She bought) a pretty plant.
3. Other:
   - a bag, I don't know.

If we look at answers to this type of question from the field testing, we find again that there is a strong effect of development, but almost no difference in AAE/MAE performance [Age, $F(5, 1002) = 11.789$, $p < .0001$; Dialect, $F(1, 1002) = .978$, $p = .323$; Age by Dialect, $F(5, 1002) = 1.471$, $p = .197$]. TD and LI groups are significantly different [Clinical Status, $F(1, 1002) = 5.866$; $p = .016$]. We find again that children who have disorders are systematically less likely to give the correct answer:

Here are the same two example children as above:

Child A	Child B
a plant	some soap

Child A has apparently either completely ignored the verb *say* or imposed an independent constraint that the lower clause must be true.

BARRIER QUESTIONS

These questions require the child to understand the concept of movement; they must know which clause a *wh*-word has moved from to answer correctly. When there is more than one clause and one or more question words, the child must be able to figure out with which verb each question word is associated. They need to be able to tell when the question is ambiguous (and there is more than one correct answer, such as, "When did she say she lost her purse?") and when there is a barrier to movement which restricts how elements can move and limits the "when" question to one correct answer (e.g., "When did she say *how* she lost her purse?")

In answering these questions, children sometimes make a very important kind of error. It can appear to be a kind of arbitrary misunderstanding, but in fact it is very systematic,

persists for a long time, and resembles grammatical structures in other languages. Such sentences are those that involve two question words, but only one question; the second question is known as an "indirect question" that we do not usually answer:

Why did you say who you saw?

This sentence asks only about why you said something, not about who you saw or why you saw them. Many languages around the world allow a construction where you answer the question in the middle (e.g., the "medial" question) and the first *wh*-word is just a clue that another word is following. It is as if one said:

What did you say how you were going to swim?

where we understand that we are to answer "how you swim." In most cases, English does not allow the possibility of answering the medial question.

De Villiers and Roeper[3] explored these questions and responses in extensive experimentation in six languages. Here is a story from their research that we presented to children in our pilot studies:

This boy was climbing in the forest one afternoon when he slipped and fell out of the tree. That night when he was taking a bath, he saw that he had bruise on his arm and he called to his dad to tell him that he had hurt himself. When did he say how he hurt himself?

Here are typical answers:

1. Short-distance responses (correct):
   - (When did he say...?) in the evening in the bath.
2. Medial answers (incorrect)
   - (. . . how fell?) he slipped from the tree
3. Long-distance responses (incorrect)
   - (When . . . fall?) in the afternoon "when he was climbing"
4. Other
   - "I don't know."

Our two case studies follow the same pattern as before:

Child A	Child B
from the tree	when he was takin' a bath

Again we find in Figure 3 that disordered children consistently do worse on these constructions across the age range, but there is no interaction between the factors [Age, $F(5, 1002) = 42.110$, $p < .0001$; Clinical Status, $F(1, 1002) = 85.857$; $p < .0001$; Age by Clinical Status, $F(5, 1002) = 1.265$, $p = .277$].

Note that what we call the medial response persists among TD children well into the 7- to 8-year range. Its persistence must have something to do with the fact that it is quite acceptable in Russian (with *how*) and German (with *what*). Thus, this interpretation is possible in some languages but not English. (The interested reader can find extensive discussion of such cases in the acquisition literature.[3,4] [Seigmuller and Weissenborn, unpublished report])

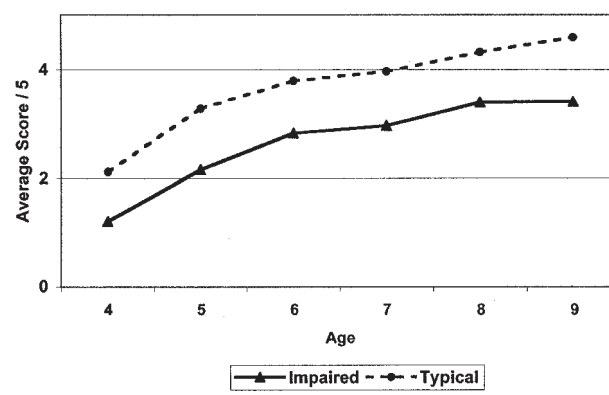


Figure 3 Barrier questions by Clinical Status.

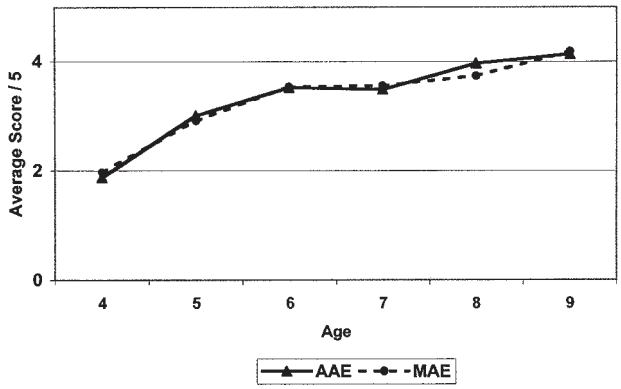


Figure 4 Barrier questions by Dialect.

As shown in Figure 4, we find in these items a strong age effect but unlike clinical status, dialect has virtually no impact [Age, $F(5, 1002) = 42.468$, $p < .0001$; Dialect, $F(1, 1002) = .073$, $p = .787$; Age by Dialect, $F(5, 1002) = .494$, $p = .781$]. This pattern makes barrier questions ideal for a linguistically fair syntax assessment.

Passive Subdomain

We turn now to another construction where some information is totally unspoken but still known to any competent speaker: passive. The passive comprehension items test children's understanding of movement and also implicit relationships; that is, hidden information that is implied by the grammar of the sentence, but not stated in words.

The most salient feature of passive constructions bears an important relation to *wh*-questions: movement is involved. The object moves to the subject position. Other properties, however, as we have outlined above, play an important role in a sentence such as:

The plant was dropped.

There is an *implied* and *disjoint* agent (that is, someone dropped the plant and the agent is different from the subject).

MOVEMENT IN PASSIVES

Control of this construction is shown if the child can reliably discern the subject and the object (Fig. 5). For example,

The dog was pulled.

The response pattern to questions of this type (Fig. 6) suggests that both LI and TD children do not always understand this distinction (as other experiments have shown as well), but the statistical comparison shows there to be a TD/LI gap [Age, $F(5, 1002) = 43.526$, $p < .0001$; Clinical Status, $F(1, 1002) = 33.949$; $p < .0001$; Age by Clinical Status, $F(5, 1002) = .719$, $p = .609$]. A separate analysis of variance showed Age and Age by Dialect effects of similar magnitude and confirmed that there was no reliable AAE/MAE difference in performance [Dialect, $F(1, 1002) = 2.686$; $p = .102$].

However, the ability to pass this item type does not demonstrate that children comprehend how the passive carries knowledge of an implicit, disjoint agent.

DISJOINT, IMPLICIT AGENT

The following illustration (Fig. 7), which is similar to those on the comprehensive language test, provides a choice between an active and an passive with a disjoint agent:

The bear was being washed.

If children prefer a result-passive reading, they do not have to assume a disjoint agent because a result-passive is much like an adjective that can be formed without any movement at all:

The bear was washed.
Compare: The bear was brown.

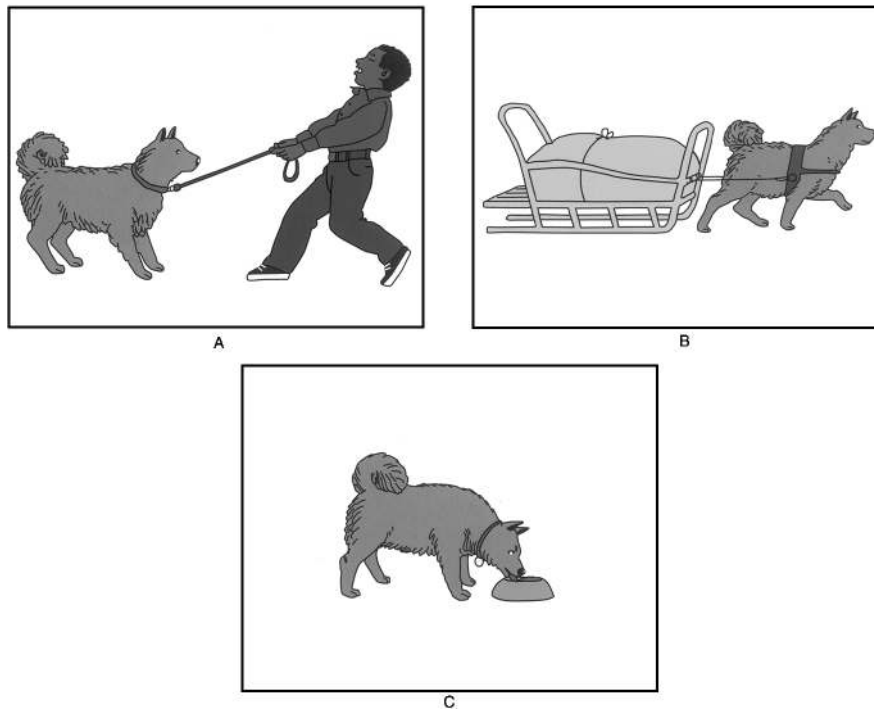


Figure 5 Basic passive item. (*Show me: The dog was pulled.*)

Therefore, this extra complication is crucial to the determination of whether the child understands the sentence to be a passive at all.

Again, the familiar pattern is present (Fig. 8), although comprehension comes somewhat later than we saw with the simple passives. The statistical comparison shows there to be a TD/LI gap with a reliable interaction by age [Age, $F(5, 1002) = 28.708$, $p < .0001$; Clinical Status, $F(1, 1002) = 29.788$; $p < .0001$; Age by Clinical Status, $F(5, 1002) = 3.935$, $p = .002$]. A separate analysis of variance showed Age and Age by Dialect effects of similar magnitude and no reliable AAE/MAE difference in performance (Dialect, $F(1, 1002) = 3.317$; $p = .07$).

DISTINGUISHING THE AGENT *BY*-PHRASE FROM A LOCATION *BY*-PHRASE (*"BY THE TREE"* AS IN *"NEAR"*)

The special role of the *by*-phrase as an agent carrier in passive can be isolated through

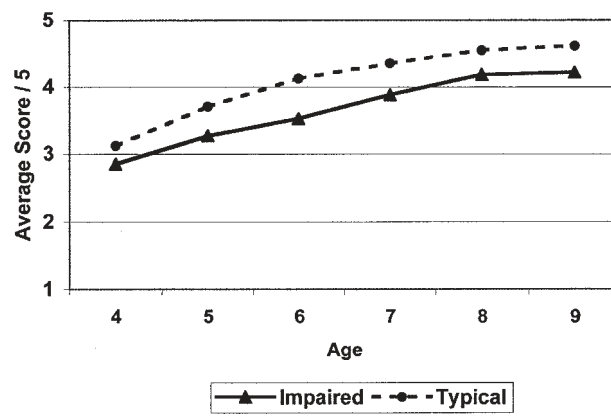


Figure 6 Basic passives by Clinical Status.

Figure 7 Complex passive item 17. (*Show me: The bear was being washed.*)

sentences that block this function. When a progressive sentence is used, the *by*-phrase can only be a locative, not an agent. Although no linguistic theory can explain very well the absence of an agent reading, it is still useful to us as a diagnostic because some children fail to recognize this:

1. The branch was dropped *by the man* (passive/agent).
2. The branch was dropping *by the man* (locative phase).

We find here that there is a significant challenge to LI children and some delay even among TD children. The statistical comparison shows no AAE/MAE difference, but a reliable difference between TD and LI children [Age, $F(5, 1002) = 9.275$, $p < .0001$; Dialect, $F(1, 1002) = .043$; $p = .835$; Age by Dialect, $F(5, 1002) = 1.681$, $p = .136$]. A separate analysis of variance confirmed Age and Age by Clinical Status effects of similar magnitude, and a strong TD/LI difference in performance (Clinical Status, $F(1, 1002) = 10.122$; $p = .002$).

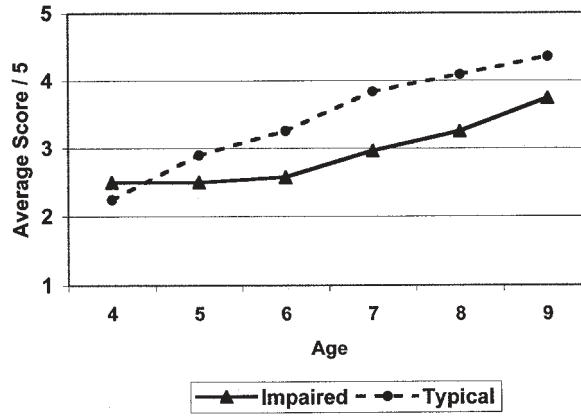


Figure 8 Complex passive items by Clinical Status.

If we now return to our two case studies, we find again that they pattern similarly:

Child A fails the simple passives 2 of 4 times, fails the complex passive 3 out of 4 times, and allowed an agent reading for a locative *by*-phrase.

Child B (18221) succeeded with simple passives 3 of 4 times and with the complex passives 2 of 4 times. Both locative *by*-phrases were correct.

Articles Subdomain: Understanding the Principles of Discourse Linking

Articles are among the most automatic forms of speech English speakers have. Yet, articles make subtle demands on children's syntax and semantics; they engage their knowledge of context and presupposition. Consider that articles constitute one of the most vexing barriers to perfect English that Asian speakers encounter. We say, "I got a cat" and are not tempted to say, "I got cat." The simple explanation for this problem is that Asian languages have no articles. The deeper question of why articles are so difficult to explain is still unclear.

Some of the Basic Uses of Articles that Can Be Tested with Children this Age

ARTICLES: DEFINITE VERSUS INDEFINITE

The most elementary bridge between sentences is simply filling in missing information, as we indicated above. We begin with the same kind of discourse link between noun phrases:

John has some berries.
I would like some (berries).
I would like the berries.

Shifting *some* to *the* depends on the recognition that it is the same berries. So also can "a dog" shift to "the dog" once it has been mentioned.

The same link is present when we apply common knowledge to justify the presupposed familiarity that *the* implies. That is, if we say:

John has some berries. The stems are still on.
(*the* stems = the stems of the berries)

We must still reconstruct a hidden connection to the berries, which again justifies *the*. It is an interesting fact observed by Schafer and deVilliers[2] that this *part/whole* connection is easily available to children in comprehension.

Indefinite articles present their own challenges. They may be either *specific* or *nonspecific*. A sentence such as,

I want to buy a book.

is ambiguous between a reading where there is a specific book that one wants and the general desire to buy a book.

Another use of the specific is when an item is known to the speaker but not the listener, which means the whole communicative situation must be conceived of. I can say,

I have a picture at home.

If I were in the room, I might get a *the* ("Look at the picture on the wall!"). However, because the listener does not know what picture is being referred to, we must use *a*.

In sum, the English article system is especially complex and, not surprisingly, it is not learned all at once by TD children. In most respects, the article system is learned alike in both MAE and AAE.

EXAMPLES OF POTENTIAL ITEMS TO ASSESS ARTICLE KNOWLEDGE

Fortunately, despite the complexity of the phenomenon, the method of eliciting answers has proven splendidly simple. We tried to avoid the concreteness that is delivered by pictures or toys, which confounds the very phenomena in which we are interested.[2,5] All we do is ask children to finish a story where the part/whole definite or other indefinites are implied.

Part/whole story:

Jack wanted to eat a coconut, but first he had to take something off it. What did he take off it?
(*The* peel, *the* shell, *the* brown thing)

Familiar-*the* story:

A snake and a bird were sitting on a rock.
They were friends. One of them flew away.
Guess which. (*The* bird)

Specific-*a* story:

I'll bet you have something sharp in your kitchen at home. What is it? (*A* fork, *a* knife)

Nonspecific-*a* story:

Chuck wants to take his teddy bear, his favorite game, and his baseball to school for show-and-tell. What can he put them all in? (*A* bag)

TYPES OF ARTICLE ERRORS

Although it is commonly assumed that children prefer specifics, we find that they substitute indefinite *a* for *the* 8 times more often than they substitute *the* for *a*. This may seem surprising from a common sense cognitive perspective, but it follows naturally from a grammatical perspective. The unmarked noun has the most abstract meaning: "John likes cake." An indefinite provides minimal marking (*a cake*), and the definite carries the most information (*I like the/that cake*).

The pattern of correct article use in TD and LI children is evident in Figure 9 [Age, $F(5, 1002) = 52.874$, $p < .0001$; Clinical Status, $F(1, 1002) = 59.893$; $p < .0001$; Age by Clinical Status, $F(5, 1002) = 6.959$, $p = .02$].

As we saw in the article by Pearson,[6] again there was virtually no difference across dialects (Fig. 10)—an ideal outcome in selecting culturally and linguistically fair items.

Case Studies Revealed

It is clear that Child A has shown worse performance on *wh*-questions, paired readings, and passive. Now we can ask an important question: Who are these children? Their backgrounds are described below.

Child A is a 5-year-old white female from the South whose parents have a high school education. She is an MAE speaker and is not receiving speech or language services.

Child B is a 4-year-old African American boy from the Midwest whose parents have a high school education. He speaks with "Some Variation from MAE" (see Ciolli and Seymour[7]) and is not receiving speech or language services.

These case studies then provide us with an ideal demonstration of the fact that deep problems can exist for MAE speakers that may not be problems for a dialect speaker. Our statistical results are borne out with individuals.

SUGGESTIONS TO HELP CHILDREN LEARN THESE KEY CONCEPTS

These examples all circle around questions where movement, sets, and truth values are written into the structure of language in ways that go beyond cognitive ability. What does this say about possible remediation? Initially, one can discuss stories of the kind used in the test with children and engage them in a deeper appreciation of them.

It will help children understand better if we take away the ambiguities that are carefully built into the test and instead create contexts that support the hidden meanings and make them evident. Some simple suggestions that address each of the missing properties follow.

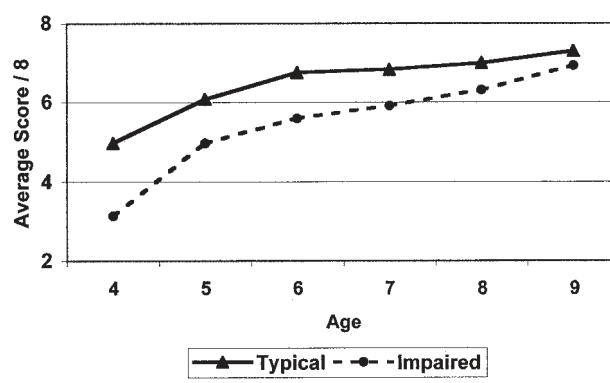


Figure 9 Article scores by Clinical Status.

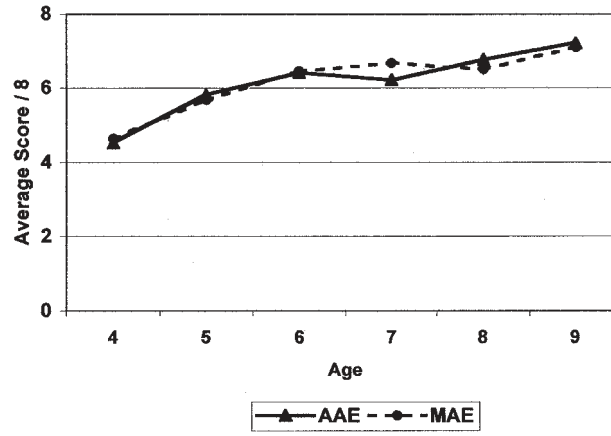


Figure 10 Article scores by Dialect.

They are all contextually clear and yet inherently indirect efforts to bring about a grammatical realization in a child.

Movement

If a child does not understand movement, we can illustrate movement first with sentences in which content is not being questioned, and so nothing needs to be supplied by the child. Thus if we say:

Milk John likes.

and then ask:

What does John like?

we have given the crucial information, not in object position at the end, but already focused in topic position at the front of the sentence.

Sets

For the child who does not understand the need to answer with a set, we can imply the multiple nature of *wh*-words through other lexical and contextual support. We can make it so that it would not make sense to answer a *wh*-question as if "who" meant "who-somebody" and not "who-everybody."

If we ask:

Who plays together?
Who shared the ice cream?
Who helped each other?

these sentences are difficult to answer sensibly with a single person. In contrast, we can imagine situations for which a single answer is impossible: For example, we can present a scene in which 20 people are lifting a rock and ask:

Who lifted the rock?

In such as example, it would not really be correct to point to just one person.

PAIRING AND EXHAUSITIVITY

To help children grasp the concept of paired exhaustive answers, one might reverse the conversational thrust to focus on exhaustivity. Suppose we have three people painting three houses and we say only:

John painted the big house.

and then ask:

Did I tell you which person painted which house?

The answer should be "no" because you only mentioned only one person and only one house. The "no" shows that the question really requires an exhaustive, paired response. In addition, one could ask the child to finish the description.

PASSIVE IMPLICIT AGENT

To help a child who has difficulty understanding the passive, the natural step is to make the hidden information explicit. First, one would

want to articulate the presence of unspecified subjects:

Someone dropped the bowl.

Second one would want to use a passive in the same situation with an explicit agent:

The bowl was dropped by someone.

Finally, one can reiterate the agentless form:

The bowl was dropped.

Thus, one could recreate the steps hidden in the formation of the passive for the child. One can show which sentences with full information are equivalent to a sentence with hidden information.

ARTICLES

One method to bring children to the awareness of part/whole relations could be to embed the relation in a conversation in which it is used correctly and see if they agree. If we show the child a picture with two cats in it, one with a brown tail and one with a red tail, and say:

Here's a cat with a long tail that is brown.
The tail is not red.
Is that right?

If the child says "yes," even though there is a red-tailed cat in the picture, then the right connection to *the* tail was clear. But if the child says "no," then we need to say that "*the* tail" picks out one particular tail and not any tail.

CONCLUSIONS

We have shown that the assessment of complex aspects of children's syntactic development between the ages of 4 and 9 can be carried out in a dialect-neutral fashion. Such assessments provide the clinician with a substantial profile of the child's language strengths and weaknesses, not just a diagnostic categorization. We have isolated certain grammatical constructions in sufficient depth that one can treat each one as a form of independent appraisal. Like ear, nose, and throat medicine, the problems in these constructions can be either connected or independent. The dialect-neutral testing allows one to achieve both a composite picture of abilities and an individualized analysis.

Our approach has allowed us to tap deep grammatical principles through unusually short and simple sentences. This allows us to minimize the impact of extraneous world knowledge, problems of parsing, or memory. We have been able to do this by using the crucial kinds of examples that have played a role in modern linguistic theory.

These materials and procedures focus on the development of aspects of language that are vital for success in early schooling and the transition to literacy. School is full of intricate questions, dialogues with missing information, and unusual connections between sentences that articles help to convey.

ACKNOWLEDGMENTS

This work was funded in part by National Institutes of Health (NIDCD) under Contract # N01 DC8-2104 and Grant # R01 DC 02171-04 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. MacWhinney B. The CHILDES (Child Language Data Exchange System) Project: Tools for Analyzing Talk, 3rd ed. Mahwah, NJ: Lawrence Erlbaum; 2000
2. Schafer RJ, de Villiers J. Imagining articles: What 'a' and 'the' can tell us about the emergence of DP. In: Howell SC, Fish SA, Keith-Lucas T, eds. BUCLD 24: Proceedings of the 24th Annual Boston University Conference on Language Development, Vol. 2. Boston, MA: Cascadilla Press; 2000:609–620
3. Roeper T, de Villiers JG. Lexical links in the wh-chain. In: Lust B, Hermon G, Kornfilt J, eds. Syntactic theory and First Language Acquisition: Cross-Linguistic Perspectives. Vol. II: Binding, Dependencies, and Learnability. Hillsdale, NJ: Lawrence Erlbaum; 1994:357–390
4. Weissenborn J, Roeper T, De Villiers J. 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 UMOP Special Edition; 1990:43–73
5. Maratsos MP. The Use of Definite and Indefinite Reference in Young Children: An Experimental Study of Semantic Acquisition. New York: Cambridge University Press; 1976
6. Pearson BZ. Theoretical and empirical bases for dialect-neutral language assessment: contributions from theoretical and applied linguistics to communication disorders. Semin Speech Lang 2004; 25:xxx–xxx
7. Ciolli L, Seymour HN. Dialect identification versus evaluation of risk in language screening. Semin Speech Lang 2004;25:33–40

Assessing Pragmatic Skills in Elicited Production

Peter de Villiers, Ph.D.[1]

ABSTRACT

In developing a test of pragmatic skills for children ages 4 to 9 years, we focused on a number of functional language skills that are important for children's success in early schooling and for the development of fluent reading and writing. They included (1) *wh*-question asking, (2) communicative role taking, (3) linking events in a cohesive narrative, and (4) articulating the mental states of the characters in a story. All of the proposed items provide specific referential support and pragmatic motivation for the forms and content to be produced by the child. The pictured materials and elicitation prompts constrain the range of appropriate utterances, so the children's productions are more easily scored than an open-ended spontaneous speech sample. All tasks described show a clear developmental trend, a clear separation between the performance of typically developing and language-impaired children, and no performance differences between African American English- and Mainstream American English-speaking children.

KEYWORDS: Communicative roles, question asking, cohesive narrative, theory of mind, mental states, perspective taking, language functions, pragmatics

Learning Outcomes: As a result of this activity, the reader will be able to identify (1) functional language skills that are included on the proposed diagnostic pragmatics probe, and (2) the developmental pattern of errors children produce on the *wh*-question asking items in the pragmatics section of the proposed assessment.

Pragmatics concerns the functional use of language in communication and discourse. Pragmatic accounts of language acquisition try to characterize children's growing communicative competence,[1–3] rather than focusing on the structural forms (syntax) or content (semantics) of their language. What does pragmatic development or communicative competence involve?

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: Peter A. de Villiers, Ph.D., Professor of Psychology, Smith College, Northampton, MA 01063. E-mail: pdevilli@smith.edu. [1]Professor of Psychology, Smith College, Northampton, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p; 2004,25,01,057,072,ftx,en;ssl00183x.

It has been suggested that seeking a single definition of pragmatics is a little like asking several gourmet pastry chefs how to bake the perfect chocolate cake.[3] Like the chefs, linguists agree on most of the basic ingredients, but they are likely to emphasize different components and so give the overall domain a different flavor. Nevertheless, research on the acquisition of pragmatic skills can usefully be organized around four major aspects of communicative competence:

1. The child's emerging conversational skills in face-to-face verbal interaction.[4] These include knowing when and how to take a turn in the conversation; how to initiate, elaborate, or terminate a topic; and how to respond to a speaker in keeping with the pragmatic constraints set by the preceding utterance (e.g., direct question forms demand answers; indirect questions [e.g., "Can you pass the salt?"] demand actions). They also include skills in detecting the presence and source of any breakdown in communication and knowing how to repair such breakdowns.
2. The developing speech acts or communicative functions of sentences in conversation.[5] For example, we use utterances to report events, to make statements (declarations) about the world, to request information or action, or to prohibit action.[6]
3. Adjusting one's language to fit the social context of the conversation in keeping with cultural conventions and social roles, whether these involve issues of politeness, formality, or the age or status of one's listener. These have been called "styles" or "registers" of speech.[7,8]
4. Finally, taking an extended turn in discourse to tell a story (narration), explain an event, give directions for how to make something or how to get somewhere, or to persuade one's listener in an argument. These are sometimes referred to as different "genres" of extended discourse.[8] They require the child to organize a series of utterances into a coherent and cohesive message.

In developing an assessment of pragmatics that would be not be biased against African American English (AAE) learners, we focused on a number of functional language skills that are important for all children's success in early schooling and for the development of fluent reading and writing. They included:

Question-answer mapping—asking the right *wh*-question to find out some specific information.

Communicative role taking—understanding the communicative perspective of others and knowing what speech acts they are producing.

Uniquely identifying referents—telling the listener who (or what) is being referred to, especially in narrating a story about several different characters.

Linking events into a cohesive narrative-expressing the temporal relationship between events.

Understanding the mental states of the characters in a story—this involves having a "theory of mind." Bruner[9] has pointed out that authentic narratives have both a "landscape of action" (the sequence of events that took place and their causal and temporal connections with each other), and a "landscape of consciousness" (the meaning of the events for the characters in terms of their emotions, desires, plans, beliefs, and states of knowledge or ignorance).

We therefore concentrated on aspects 2 and 4 of the components of communicative competence listed above. We did this for two primary reasons. First, style or register adjustments of speech for reasons of formality, status, or age (aspect 3 above) vary with cultural conventions, and probably vary with cultural groups that speak different dialects of English, so they do not lend themselves to a dialect-neutral assessment of pragmatic development. Second, interactive conversational skills (aspect 1) are best assessed in ongoing conversation or language sampling rather than in a formal, picture-based test.

There are certain key features of all of the elicitation materials and procedures in the pragmatics tasks that follow. First, they provide specific referential support and pragmatic motivation for the language forms and content to be produced by the child, so they greatly increase the likelihood that those forms and

functions will be sampled in the assessment. Second, the pictured materials and the elicitation prompts constrain the range of appropriate utterances, so the children's productions are much more easily scored than a more open-ended spontaneous speech sample. However, the procedures retain a considerable degree of communicative naturalness rather than resorting to unnatural imitation procedures to elicit the forms. All of the subdomains test the interaction between syntactic and semantic forms with specific pragmatic functions, the inseparable interaction among form, content, and function in language acquisition described by Bloom and Lahey.[10] Assessment of pragmatic skills cannot be divorced from the syntactic forms and semantic meanings that are required for those functions of language. Finally, all of the materials are picture-based, so they require minimal technology and can be administered and scored by a single clinician interacting with the child.

ASKING *WH*-QUESTIONS

Young children must master a variety of *wh*-question forms in English that request different kinds of information from the listener—specification of objects (*what*), persons (*who*), locations (*where*), reasons and causes (*why*), instruments or manners of action (*how*), or times (*when*).[11,12] We developed a set of probes to elicit what, who, where, why, and how questions. The test also elicits a more complex double *wh*-question form ("Who is eating what?" or "Who is eating which food?") that indicates whether the child understands the distributive set properties of complex *wh*-questions. (See Roeper[13] for a description of the important semantic and syntactic properties of double *wh*-questions.)

In the elicitation procedure, the child is shown a picture with something missing from it. The area of the missing element of the picture is a blank space surrounded by a dotted line.

The child has to ask "the right question" to find out "what is happening in the picture." The missing elements of the pictures include objects, people, locations, tools, and causes of emotion—so *what*, *who*, *where*, *how*, and *why* questions are naturally motivated by the pictures. The idea of the game is communicated in two warm-up items (a "what" and a "who" question) in which the tester uses a great deal of prompting to introduce the child to the game so that they come to ask questions rather than just guess at the answer. An item of this type is shown in Figure 1. A girl is shown holding a paintbrush and working on some object (an irregular space surrounded by a dotted line). The tester prompts the child with the following: "The girl is painting *something*. You need to find out *what* she is making. Ask me the *right* question, and I'll show you the answer." (Italicized words are emphasized). If the child does not ask an appropriate *wh*-question, the tester continues to prompt, modeling a correct what question: "Ask me, *What is the girl painting*? You say it. What . . .?" When the child asks an appropriate question, the tester turns the page of the stimulus book and shows the child a completed picture of the girl painting a chair.

After the warm-up questions there follow 9 test items, covering 5 different *wh*-question forms as specified above, plus one double *wh*-question. The amount of prompting provided by the tester varies across the items to create differing amounts of scaffolding for the child, and hence, differentially difficult items in terms of pragmatic skill. Thus, for the first four items, the tester begins by giving a semantic domain prompt that tells the child the general semantic category of the desired information: e.g., "The boy is calling *somebody*. Ask me the *right* question, and I'll show you the answer." If the child does not produce an appropriate who question, the secondary prompt is to provide the *wh*-word: "*Ask me a who question*. Who . . .?" Thus, the maximum amount of prompting for these items is providing the correct *wh*-word to use.

For the next four items, there is no semantic domain prompt at the beginning, and the tester goes straight to "Ask me the right question, and I'll show you the answer." Thus, the child has to use the pictured event alone to determine what question is needed. On these trials, if the child does not produce an appropriate *wh*-question, the secondary prompt for the tester is to specify the semantic domain of the missing information. Thus, the maximum amount of prompting for these items is for the

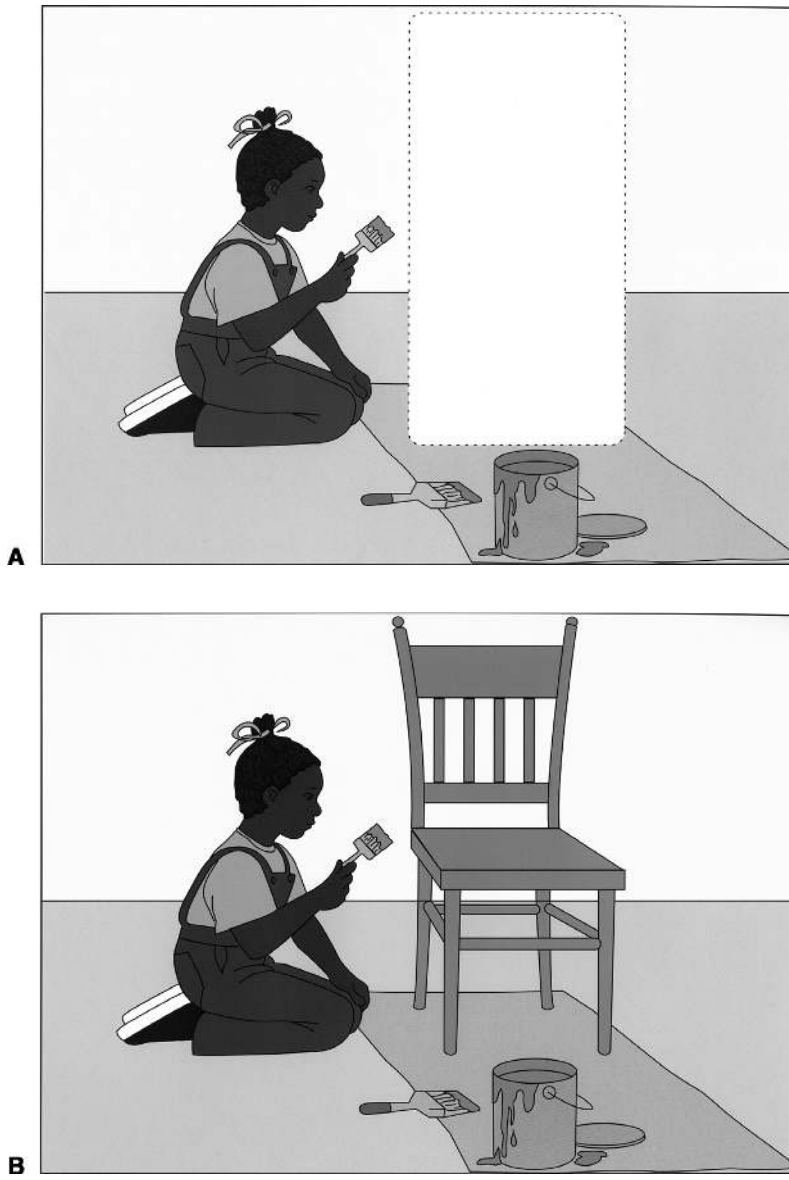


Figure 1 (A, B) Pragmatics trial item (question asking).

semantic domain of the desired question to be provided to the child.

The children's responses are scored as whether the child produces a semantically and pragmatically appropriate *wh*-question for each item. The exact *wh*-question form produced can vary in some cases and still be acceptable for that item. For example, if the target question were "Why is the girl sad?" the responses "What is she sad about?" and "What is she sad for?" are also correct. Similarly, the syntactic form was allowed to vary to accommodate dialect variation in morphosyntax. Thus, "What she paintin'?" is as appropriate in pragmatic terms as "What is she painting?" Thus, the children were given one point for each item for which they produced a pragmatically appropriate *wh*-question following all levels of prompting that they received. (For purposes of diagnostic assessment in the current test, the level of prompting needed for a particular item for a child was not differentially weighted

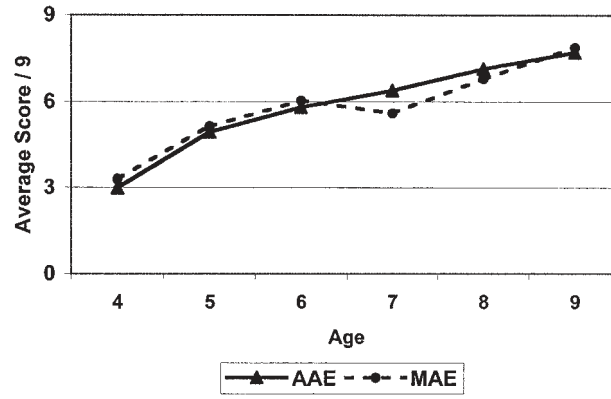


Figure 2 Question asking by Dialect.

in the scoring, but this information of how much scaffolding each child needed before he or she produced an appropriate question is available to the clinician.)

The first question to be asked is whether these pragmatic probe items were biased against AAE-speaking children? Figure 2 indicates that the task produced strong developmental data, with substantial growth in performance between the ages of 4 and 9 years. However, there was no significant difference in performance between AAE and MAE speakers at any age. [Age, $F(5, 1002) = 50.876$, $p < 0.0001$; Dialect: $F(1, 1002) = 0.034$, not significant [n.s.]; Age by Dialect, $F(5, 1002) = 0.556$, n.s.]

On the other hand, the question-asking task distinguished clearly between typically developing children and language-impaired children across the entire range of ages (see Fig. 3). (Age, $F(5, 1002) = 58.237$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 79.612$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 0.338$, n.s.)

The data from each *wh*-question item looked very much like the overall data summed over all of the items—there were no differences between AAE and MAE speakers, but strong separation between the graphs for typically developing and language-impaired children.

Table 1 shows sample performances from two 5-year-old children, two 6-year-old children, and two 8-year-old children on the question-asking subdomain. All six children are speakers of AAE. At each age, one child is typically developing and the other is language-impaired. The responses that were coded as incorrect are underlined. The table illustrates both the development between ages 5 and 8 years, and the difference between typically developing and language-impaired children.

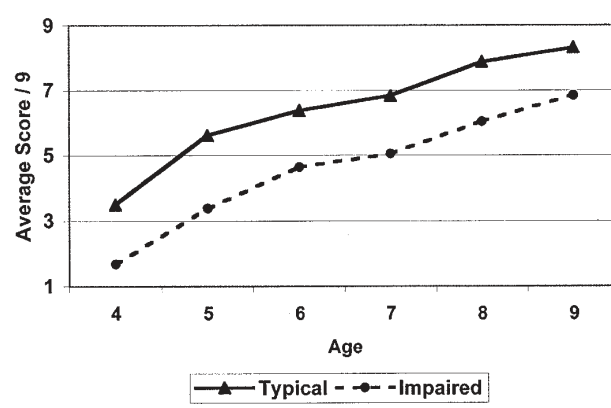


Figure 3 Question asking by Clinical Status.

Table 1 Example Verbatim Responses on the Question Asking Subtest

Target	5-Year-Old Typical AAE	5-Year-Old Disordered AAE
Who	What is the nurse feeding?	No response
Where	Where did she go swimming?	She makin' a pool.
Why	What is the girl mad about?	What the girl.
How	How is the girl fixin' that?	She's fixing his bike.
What	What is the woman eating?	What some meat.
Who	Who is riding the bike?	What the boy.
Where	Where is that boy going?	The boy is running to the ice cream.
Why	What happened?	What?
Who-what	What is they eating?	No response
Target	**6-Year-Old Typical AAE**	**6-Year-Old Disordered AAE**
Who	Who is the nurse feeding?	Who is that feeding him?
Where	Where did the girl swim?	She jumped in the water.
Why	What is the girl mad for?	She mad at the table.
How	What is the girl fixing?	She is fixing the toy.
What	What is the girl eating?	Who's eating?
Who	Who is riding the bike?	A boy riding on the bike.
Where	Where is the boy running?	Who's running?
Why	Why is the boy crying?	He dropped his ice cream.
Who-what	What are the people eating?	Who's eatin'?
Target	**8-Year-Old Typical AAE**	**8-Year-Old Disordered AAE**
Who	Who is the nurse feeding?	Who is she feeding?
Where	Where did the girl go swimming?	What something she swim in?
Why	Why is the girl mad?	Who is she mad at?
How	How is the girl fixing the toy?	What's she holding on her hand?
What	What is the woman eating?	What her mom eating from her two fingers?
Who	Who is riding the bike?	Something riding a bicycle.
Where	What is the boy running to?	Where is his house?
Why	Why is the boy crying?	Was he crying?
Who-what	Who is eating what food?	How was they was eating?

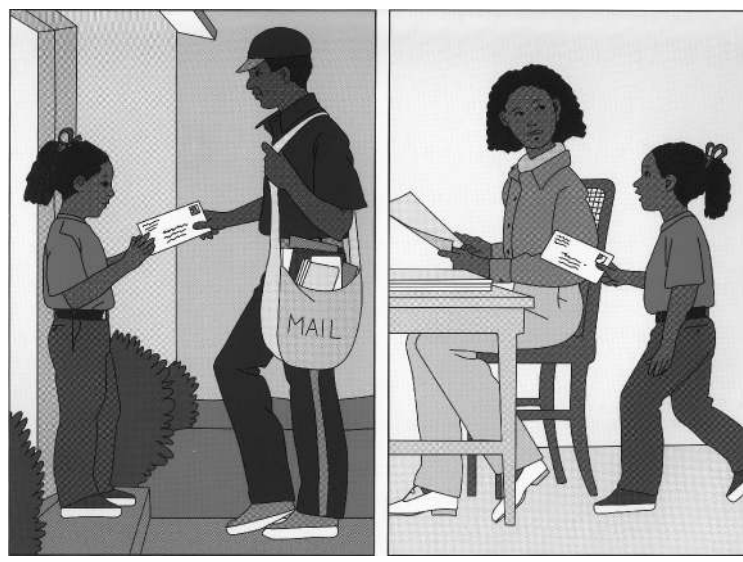
*Underlined responses were scored as incorrect and did not receive any credit.
(AAE, African American English.)

The dominant "error" for the typically developing 4-year-old children was a failure to even ask a question, with children making a guess at what the answer might be. In the 5- and 6-year-old children, the most frequent errors were asking the wrong *wh*-question for the information needed or asking an all-purpose question that was too vague, such as "What is he (or she) doing?" or "What is it?" The older children tended to get all or almost all of the single *wh*-questions correct, but they were often still unable to produce a correct double *wh*-question. A similar pattern of errors was seen in the language-impaired children, but each of these developmental errors tended to persist for longer, with even the older children still asking the wrong *wh*-question or asking all-purpose, nonspecific questions.

COMMUNICATIVE ROLE TAKING

Children's ability to take the perspective of another speaker and to understand what speech act they were producing was tested in a communicative role-taking task (Fig. 4). For each trial the child was shown a sequence of two pictures. In the first picture a character either participated in or observed an event. For example, the tester might point to the picture and say: "Look at what is happening here." A

A



B

Figure 4 (A, B) Communicative role taking (communication roles).

second picture is then revealed in which the character from the first picture is gesturing and clearly saying something to another person, or in which the character from the first picture was clearly being spoken to by the newly introduced person. Depending on the nature of the sequence of events, the child was asked by the tester what the speaking character in the second picture was "telling," "asking," or "saying to" the other person in the second picture. The pictured events and the communication verb used by the tester served to constrain the type of speech act that the child being tested should produce.

An example of such an item might show a little girl at her door taking a letter from the mail carrier in the first picture. Then in the second picture she is shown handing the letter

to her mom. The girl appears to be talking, and the mother is in a posture of listening.

The tester prompts the child with the questions: "What is the girl *telling* her mother?" The use of "tell" in the prompt constrains an appropriate response from the child to be a statement, either in direct or indirect speech, although the specific form or content of the statement can vary somewhat given the pictured event. In an example such as this one, "Here's some mail" or "that the mail came" are both fine answers. However, a question form such as "Can I go outside and play?" violates the pragmatic constraint introduced by the prompt.

Similarly, direct or indirect question forms can be elicited by an item in which one character asks the other something. If, for example, a boy is shown in one picture looking under his bed for something, it could set the scene for a questioning sequence. In the second picture, he is turning to his mother who is shown standing next to him and the boy is asking her something. The tester might prompt with "What is the boy *asking* his mother?" The use of "ask" in the prompt constrains an appropriate response from the child to be a question or request, either in direct or indirect speech, although the specific form or content of the question can vary. So in this example, "Do you know where my ⟨baseball⟩ is?" or an indirect form such as "if she has seen his ⟨toy⟩" are pragmatically appropriate answers. The child can fill in what kind of thing he is looking for. However, in a scenario like this a declarative form, such as "there is something under my bed" or "I am trying to find my ball" would not be an appropriate response.

This procedure thus tests both the children's ability to take the communicative role of the speaking character in the picture sequence and seeing the events from their point of view, and also their sensitivity to the pragmatic constraints placed on their response by the prompt produced by the tester.

We developed a probe containing four items, one reporting an observed event ("telling"), two requesting an object or action ("asking"), and one prohibiting an action or scolding the person who did it.

Table 2 gives typical verbatim responses to similar items from four 4- and 6-year-old AAE-speaking children. At each age one child is typically developing and the other is language impaired.

Table 2 Example Verbatim Responses on the Communicative Role Taking Subtest

Prompting Verb	4-Year-Old Typically Developing AAE
Tell	His big sister felled off her bike.
Ask	Can I play baseball?
Ask	Can I have a piece of cake?
Say	You don't feed the dog. That's his own feed.
Prompting Verb	**4-Year-Old Language-Impaired AAE**
Tell	Her bleedin'.
Ask	Him carry something.
Ask	Her say look at the cake.
Say	He feedin' the dog.
Prompting Verb	**6-Year-Old Typically Developing AAE**
Tell	That his sister got hurt.
Ask	Can he go outside and play with his brother.
Ask	Can I get a piece of that cake?
Say	Don't give the dog none of your food.
Prompting Verb	**6-Year-Old Language-Impaired AAE**
Tell	She got an owie, a sore.
Ask	I got a bat and a glove.
Ask	She can eat cake yet?
Say	No.

*Underlined responses were scored as incorrect and did not receive any credit.
AAE, African American English.

Figure 5 shows that this is a developmentally sensitive assessment of children's understanding of communicative roles and speech acts that is not biased against speakers of AAE. The developmental growth curves of performance for the two dialect groups, AAE and MAE, fall directly on top of each other. [Age, $F(5, 1002) = 46.901$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.025$, $p = 0.875$; Age by Dialect, $F(5, 1002) = 0.620$, $p = 0.685$.]

The task also is strongly discriminating of language impairment. Figure 6 indicates that especially at the younger ages (between 4 and 6 years) there is a clear separation in

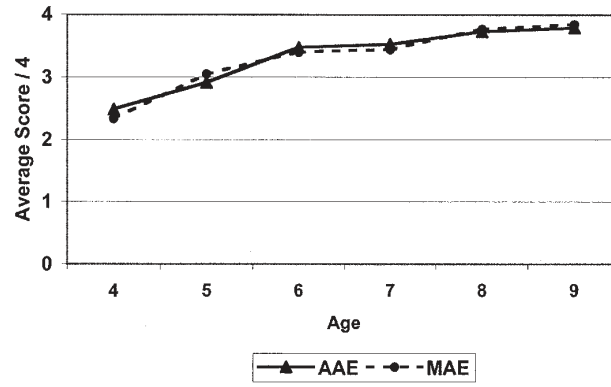


Figure 5 Communicative roles by Dialect. (AAE, African American English; MAE, Mainstream American English.)

performance between typically developing and language-impaired children. However, in this task, unlike the question-asking task, the difference between the two groups seems to narrow with age, and the language-impaired children seem to catch up with their typically developing peers by ages 8 and 9. [Age, $F(5, 1002) = 53.549$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 41.486$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 1.590$, $p = 0.160$.]

NARRATIVE

Narrative is a fundamental way in which we encode and make sense of our experiences and communicate them to others.[9] It is also the first genre of reading and writing that children do, so the acquisition of good narrative skills is crucial for early literacy development.[14]

Because narrators are free to choose different perspectives on events, there is no single type of story for any given event, and the telling of several kinds of stories has been studied.[15,16] However, psycholinguists argue that well-formed versions of each of these types of stories have two things in common. The first is thematic *coherence* on the macro-level of the overall structure or organization of the events. The second is the linguistic *cohesion* of the discourse at the micro-level of referents and clauses. Across languages and dialects, and across different types of story structures and themes, there is a common developmental pattern toward increasing coherence and cohesion in children's story telling.[17,18]

Given that there are data to suggest that AAE-speaking children produce a wider range of different story structures when given an

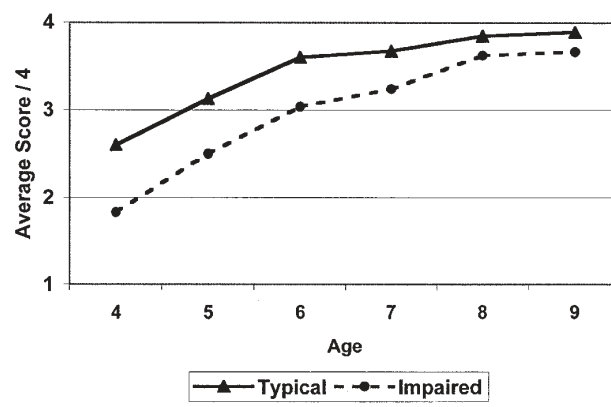


Figure 6 Communicative roles by Clinical Status.

open-ended story-telling task,[15,19–21] in developing a dialect-sensitive narrative assessment we decided to focus on the developing *cohesion* in children's stories based on a more structured picture sequence. It was reasoned that the development of linguistic cohesion would be more dialect neutral than the overall organization or story grammar of narratives, given that fundamental discourse cohesion devices are required in the whole range of different story types to link together and relate to each other the characters and events.[22]

The narrative assessment in the current test therefore concentrates on two features of children's mastery of linguistic cohesion in their production of stories: the contrastive specification of referents (telling the listener who I am referring to as each action and event is described), and the linking together of the events of the story in time. Each of these features has been shown to be revealing of developmental growth and language delay in children.[18,23–26]

In addition, several authors have argued that authentic stories do not only relate the sequence of actions and events (the "outside view"[27] or "landscape of action"[9]). They must also make reference to the meaning of the events for the protagonists (the "inside view"[27] or "landscape of consciousness"[9]) of the narrative. Painting the landscape of consciousness in a story requires the story-teller to have developed a "theory of mind"[28–30] in which the narrator uses language about mental states to explain the characters' actions and reactions in terms of their emotions, desires and cognitions—what they think or know (or don't know) at different points in the story.[9] Children's growing understanding and ability to express the mental states of the characters can be studied effectively in well-designed picture sequence narratives.

The narrative elicitation picture sequence therefore has three fundamental features that provide strong pragmatic motivation for the expression of these aspects of linguistic cohesion and references to mental states. First, there are two characters of the same gender interacting throughout the scenario who need to be referred to contrastively (and pronouns alone will not be sufficient). Second, there are important temporal relationships between the events in the pictures (both within a picture and between pictures) that must be expressed in a fluent, cohesive narrative. Finally, the pictured scenario is based on standard tests of theory of mind reasoning in children in which a desired object is moved from one location to another without the major protagonist observing the change. A "thought balloon" is used in the picture sequence to depict the mental state of the character, and the child is asked at the end of the story why the character goes to look for the object in the wrong place (i.e., why he has a false belief about the location of the object).

The picture sequence is presented on the page of a tented stimulus picture book facing the child and away from the examiner so the child is reminded that the examiner cannot see the story pictures. The child is told to look carefully at each picture to see what happens in the story and then to start at the beginning and tell the whole story to the examiner. While the child tells the story, the examiner notes on a checklist whether the child contrasts the two boys so the listener can tell them apart and what kinds of temporal linking expressions the child uses. The target narrative cohesion elements are therefore scored online while the child is telling the story. Checking online scoring against recorded and transcribed children's narratives showed that the reference contrasting and temporal marking could be scored online with sufficient reliability (with 87.5% accuracy in a sample of more than 80 narratives).

After the child has told the story, the examiner points back to the picture in the sequence that shows something like a little sister reentering the kitchen with a thought balloon depicted to show that she is thinking about the cake she and the older sister have made together. The child is asked to tell what is happening in that picture again. The child's response is scored for whether it simply refers to the character's actions (e.g., "The girl is coming into the room."), whether it refers to her intentions or desires (e.g., "She wants her cake." or "She is coming back to get her cake."), or whether it refers to the cognitive state of the character (e.g., "She is thinking about her cake.").

Then the examiner points to the final picture in the sequence, in which the little sister is looking for the cake in a cabinet although her big sister moved it into the refrigerator while the little girl was out of the room. The examiner says to the child: "The little girl is looking for the cake in the cabinet. Why is she looking *there*?" (with the emphasis on the "there"). This is a standard theory of mind test for the child's ability to explain a person's behavior in accordance with a false belief.[31]

Table 3 shows verbatim narratives elicited from two typically developing 4-year-old children and two typically developing 6-year-old children to illustrate the developmental stages observed in the different narratives told by the children. Younger children typically did not specify the characters clearly for the listener, referring to both boys in the same way throughout, either as "the boy" or by the indiscriminate use of the pronoun "he." In keeping with the research on temporal referencing in developing narration, the children went from using no temporal connectors between events or stringing the pictured events together with "and," to using sequencers such as "then" or "and then," to using more and more temporal adverbs and adverbial clauses of time ("while," "when," or "after").[18,23] In describing the thought balloon picture, children went from simply describing the boy's actions (the dominant response at age 4) to talking about his intention or desires, and finally to specifying his cognitive state. On the theory of mind question about the last scene, there was a developmental shift from being unable to give any explanation, to explaining why the character was looking for the train (e.g., "Because he wants his train."), to providing an adequate explanation for why he was looking in the wrong place (e.g., "He thinks it is under the bed." or "Because he put it there."). This too is in keeping with the established developmental sequence in theory of mind.[28,32]

Contrasting the characters was scored as present (1 point) or absent (0 points), but the use of temporal expressions and the child's references to mental states were scored according to developmental sophistication. Thus, the use of sequencers only received 1 point but adverbial clauses of time received 2 points. Similarly, referring to intentions or desires received 1 point, but reference to cognitive states (e.g., "thinking," "wondering," "knowing") in describing the thought balloon picture or explaining why the character was looking in the wrong

Table 3 Narrative Samples Showing Typical Development between Age 4 and 7 Years

Typically Developing MAE; Age 4 Years, 2 Months

I want my train. I'm gonna hide the train from him. I'm gonna play out of the toy box. I'm gonna find that train. Bring that train.

Typically Developing MAE; Age 4 Years, 9 Months

He was looking for the choo choo train because the other boy was playin'. And then... and then he said, ''I want that choo choo train back'', and umm... he put it in his toy box. And then he came back to find it and he looked under the bed and it wasn't there.

Typically Developing MAE; Age 6 Years, 4 Months

The big boy came into the little boy's room and took away the little boy's train. Then he hid it under the boy's bed where he couldn't get it. Then the little boy... when he left... he got out his train and put it in the toy box while the big boy was eating. Then the big boy thought about the train and he went under the bed to go see it but it wasn't there.

Typically Developing AAE; Age 6 Years, 3 Months

The little brother was trying to get his toy from the big brother. And the big brother hiding his toy under the bed. When he is eating his sandwich, the little boy go and get it and put it inside of his toy box. When his big brother walk in, he think about the train and he look under his bed for it.

MAE, Mainstream American English; AAE, African American English.

Table 4 Scored Narrative Features for Four AAE-Speaking Children

Feature	4-Year-Old Typically Developing AAE	Points
Reference contrast	Yes	1
Temporal links	Sequencer	1
Thought balloon	He dreamed that his train was under the bed.	2
Why looking there?	Because he wanted it.	1
	Narrative score	5
Feature	**4-Year-Old Language-Impaired AAE**	**Points**
Reference contrast	None	0
Temporal links	None	0
Thought balloon	The boy take that for him.	0
Why looking there?	Cause he got find the train.	1
	Narrative score	1
Feature	**6-Year-Old Typically Developing AAE**	**Points**
Reference contrast	Yes	1
Temporal links	Adverbial clause	2
Thought balloon	The big brother is thinking about the train and he going back to his room.	2
Why looking there?	He think it's there	2
	Narrative score	7
Feature	**6-Year-Old Typically Developing AAE**	**Points**
Reference contrast	None	0
Temporal links	Sequencer	1
Thought balloon	The boy can't find the choo choo train	0
Why looking there?	Because he can't find it	1
	Narrative score	2

AAE, African American English.

place received 2 points. Thus, the narrative subdomain was scored out of a possible 7 points (Table 4).

Figures 7 and 8 show that there was strong developmental growth in these features of the children's narratives, but there was no difference between the speakers of MAE and AAE at any age. [Age, $F(5, 1002) = 58.152$, $p < .0001$; Dialect, $F(1, 1002) = 1.341$, $p = 0.247$; Age by Dialect, $F(5, 1002) = 1.884$, $p = 0.094$.] However, the narrative elicitation was clearly discriminating between the typically developing

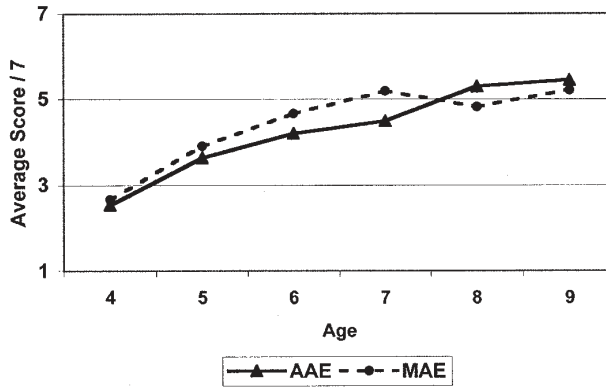


Figure 7 Short narratives by Dialect. (AAE, African American English; MAE, Mainstream American English.)

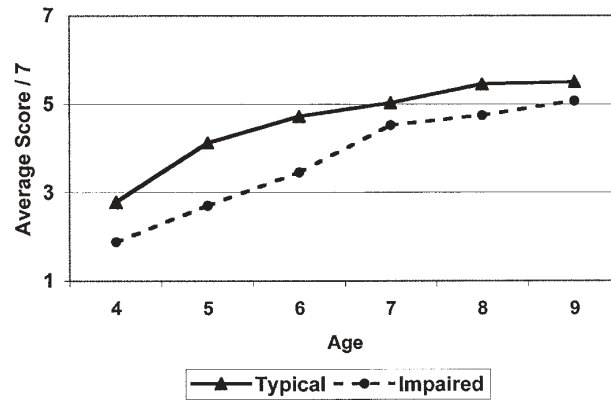


Figure 8 Short narratives by Clinical Status.

children and the language-impaired group. [Age, $F(5, 1002) = 64.390$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 50.861$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 2.056$, $p = 0.069$.] The language-impaired children showed persistent delay in all three of the features of narrative being assessed: reference contrasting, temporal expressions, and theory of mind.

THE PRAGMATICS DOMAIN

To obtain a pragmatics profile, the scores for the three pragmatics subdomains (question asking, communicative role taking, and narration) were combined into a Pragmatics domain score. Because there were only 4 items on the communicative role taking subdomain, those items were each worth 2 points, so that that subdomain contributed about the same proportion to the overall score as the other two subdomains (9 items for question asking and 7 points for narration). This produced a Pragmatics domain score with a maximum of 24.

Figures 9 and 10 show that this domain score has the properties that are necessary for an effective, unbiased assessment of dialect-speaking children. There is strong developmental growth in the measure across the ages of 4 to 9 years, but the growth functions for the MAE- and AAE-speaking children fall right on top of each other, so there is no bias against either of the dialects: Age, $F(5, 1002) = 88.732$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.050$, $p = 0.823$; Age by Dialect, $F(5, 1002) = 0.443$, $p = 0.818$. Second, there is a clear separation at all of the ages between the performance of the typically developing children and the children who were identified by the clinicians as being language impaired: Age, $F(5, 1002) = 108.509$, $p < .0001$; Clinical Status,

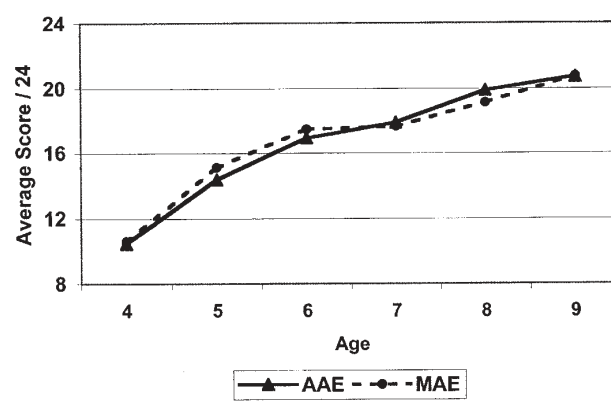


Figure 9 Pragmatics domain by Dialect. (AAE, African American English; MAE, Mainstream American English.)

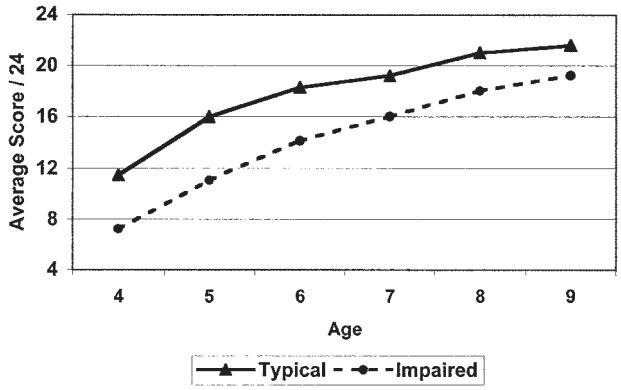


Figure 10 Pragmatics domain by Clinical Status.

$F(1, 1002) = 110.714$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 1.357$, $p = 0.237$.

CONCLUSION

These elicited production procedures provide a dialect-neutral assessment of pragmatic skills in several important subdomains that are crucial for young children's success in school and their early literacy development: asking the right question to obtain specific information, taking the perspective of a communicator and understanding what speech act they are producing, and producing a cohesive narrative that clearly identifies the protagonists for the listener, expresses the time relationships between events, and makes reference to essential features of the mental states of the characters.

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. Bates E. Language and Context: The Acquisition of Pragmatics. New York: Academic Press; 1976
2. Hymes D. On communicative competence. In: Pride JB, Holmes J, eds. Sociolinguistics. New York: Penguin; 1972:269–293
3. Snyder LS, Silverstein J. Pragmatics and child language disorders. In: Schiefelbusch RL, Lloyd LL, eds. Language Perspectives: Acquisition, Retardation, and Intervention, 2nd ed. Austin, TX: ProEd; 1988:189–222
4. McTear M. Children's Conversation. Oxford, UK: Blackwell; 1985
5. Dore J. Conditions for the acquisition of speech acts. In: Markova I, ed. The Social Context of Language. Chichester, UK: Wiley; 1978:87–112

6. Searle JR. Speech Acts: An Essay in the Philosophy of Language. London: Cambridge University Press; 1969
7. Andersen E. Speaking with Style: The Socio-Linguistic Skill of Children. London: Routledge; 1990
8. Clark E. First Language Acquisition. Cambridge: Cambridge University Press; 2003
9. Bruner J. Actual Minds, Possible Worlds. Cambridge, MA: Harvard University Press; 1986
10. Bloom L, Lahey M. Language Development and Language Disorders. New York: John Wiley; 1978
11. de Villiers JG, de Villiers PA. Acquisition of English. In: Slobin D, ed. The Crosslinguistic Study of Language Acquisition: Vol. 1. The Data. Hillsdale, NJ: Lawrence Erlbaum; 1985:27–140
12. James S, Seelbach M. The pragmatic functions of children's questions. J Speech Hear Res 1982;25: 2–11
13. Roeper T. Diagnosing language variations: underlying principles for syntactic assessment. Semin Speech Lang 2004;25:xxx–xxx
14. Tabors P, Roach KA, Snow CE. Home language and literacy environment: final results. In: Dickinson DK, Tabors P, eds. Beginning Literacy with Language. Baltimore: Paul H. Brookes; 2001:111–138
15. Champion T. Understanding Storytelling Among African American Children: A Journey from Africa to America. Mahwah, NJ: Lawrence Erlbaum; 2002
16. Toolan M. Narrative: A Critical Linguistic Introduction. London: Routledge; 1988
17. Bamberg M. The Acquisition of Narratives: Learning to Use Language. Berlin: Mouton; 1987
18. Berman R, Slobin D. Relating Events in Narrative: A Cross-Linguistic Developmental Study. Hillsdale, NJ: Lawrence Erlbaum; 1994
19. Gee JP. Two styles of narrative construction and their linguistic and educational implications. J Educ 1989;171:97–115
20. Heath SB. Ways with Words: Language, Life, and Work in Communities and Classrooms. New York: Cambridge University Press; 1983
21. Michaels S. Sharing time: children's narrative styles and differential access to literacy. Lang Soc 1981; 10:423–442
22. Halliday MAK, Hassan R. Cohesion in English. London: Longman; 1976
23. Berman R. On the ability to relate events in narrative. Discourse Process 1988;11:469–497
24. de Villiers PA. English literacy development in deaf children: directions for research and intervention. In: Miller J, ed. Research in Child Language Disorders: A Decade of Progress. Austin, TX: ProEd; 1991:349–378
25. Liles B. Cohesion in the narratives of normal and language disordered children. J Speech Hear Res 1985;28:123–133
26. Liles B. Episode organization and cohesive conjunctives in narratives of children with and without language disorder. J Speech Hear Res 1987;30: 185–196
27. Bruce B. A new point of view on children's stories. In: Anderson RD, Osborn J, Tierney RJ, eds. Learning to Read in American Schools. Hillsdale, NJ: Lawrence Erlbaum; 1984
28. Wellman HM. The Child's Theory of Mind. Cambridge, MA: The MIT Press; 1990
29. Perner J. Understanding the Representational Mind. Cambridge, MA: The MIT Press; 1991
30. Astington J. The Child's Discovery of the Mind. Cambridge, MA: Harvard University Press; 1993
31. Bartsch K, Wellman H. Young children's attribution of action to beliefs and desires. Child Dev 1989;60:946–964
32. Wellman H, Bartsch K. Before belief: children's early psychological theory. In: Lewis C, Mitchell P, eds. Children's Early Understanding of Mind: Origins and Development. Hillsdale, NJ: Lawrence Erlbaum; 1994:331–354

Cultural and Linguistic Fairness in the Assessment of Semantics

Jill G. de Villiers, Ph.D.[1]

ABSTRACT

In creating a semantic assessment that will be linguistically and culturally fair, it is important to avoid the bias of acquired vocabulary tests. This article describes techniques to assess children's processing of new words, their lexical organization and retrieval, which may be more significant than the number of words known. Special properties of the quantifier *every* also give an idea of the child's level of semantic ability. In addition, we examine how the child's performance on each item-type helps the speech-language pathologist determine the sources of children's problems in this domain.

KEYWORDS: Lexical organization, fast mapping (novel verbs), transitive structures, transfer structures, complement structures, acquired vocabulary, quantifiers, argument structure, learning from context

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) how the semantic items address the issues of linguistic and cultural bias, and (2) how the proposed semantic tasks differ from semantic tasks on traditional language tests.

KEY CONCEPTS

Semantics is central to language acquisition; it is about the expression and understanding of meaning. It is usually construed as being about the mental dictionary of words, or lexicon, that the child has to acquire. Word learning is crucial for a language user, and it needs to be fast and efficient. Not only that, but once words are learned, they need to be stored in an organized way so retrieval is efficient. However, semantics is broader than the lexicon alone in that it also includes the special properties of how words such as quantifiers (*every*, *all*, *some*, *none*) interact in sentences.

Point 1: Words Depend on Input; Input Depends on Culture

How does the assessment of semantics impact on culture- and dialect-fairness? First, word

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: Jill G. de Villiers, Ph.D., Professor of Psychology and Philosophy, Smith College, Northampton, MA 01063. E-mail: jdevil@smith.edu. [1]Professor of Psychology and Philosophy, Smith College, Northampton, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25;01,073,090,ftx,en;ssl00185x.

learning, more than any other domain, is highly subject to frequency in the input to the child. After all, words are arbitrary symbols, so there is no way the child can use innate knowledge to guess what a chair might be called. But the consequence is that children growing up in different households, in different conditions of input, will develop vocabularies that differ from each other, and grow possibly at different speeds. Some families may spend a lot of time reading picture books exposing the child to words outside of their usual experience, such as *hippo*, or *octopus*, or *volcano*. Another family may spend more time in discourse about family events, talking about relationships, emotions, or activities, which would expose the child to words such as *great-grandma, college, promotion*, and *pride*. Yet another family may engage their young children in chores such as shelling peas or shucking corn, in which case the child may develop vocabulary of a rich sort about a limited domain such as the garden: *daddy long-legs, husk, pod,* and *compost*. We were concerned that many previous studies had suggested that African American children reveal a more limited vocabulary on vocabulary tests, usually picture-based choice tests. These tests in general have been standardized on a sample representing very few such children.[1] It is a fact of American life that although African American children may go to some of the same schools and live in some of the same neighborhoods as their white counterparts, there is nevertheless a rich culture that is distinctive of African American families.[2,3] Participation in that different culture is likely to give the African American English (AAE)-speaking child exposure not only to a different vocabulary but a different emphasis on which words are central to childhood. The picture-based tests are often heavily biased toward nouns, because verbs are harder to capture in a still picture. Yet Blake[4] has argued that verbs may form a more important part of the vocabulary for AAE-speaking children, just as they do for some languages such as Korean.[5] Thus, the reduced vocabulary seen on standardized tests may be a biased estimate of AAE-speaking children's competence in semantics.

The outcome on acquired vocabulary tests may tell us that a particular child is less likely than another to be prepared for schooling, especially if school requires the same type of vocabulary as the test. However, being culturally different is not the same as having a language disorder, as the American Speech and Hearing Association has been at pains to point out for several decades.[6,7]

Point 2: Words Are Fast Mapped from Rich Contexts

Parents and teachers working with a very young child may use pointing and single word use to name a new object, such as *train*. However, once the child gets to be a toddler or older, much vocabulary is learned "on the fly" from conversational context. A new word is embedded in a sentence, and the child must identify the new word and use cues from the sentence to home in on what part of speech it is. Knowing that will allow the child to scan the context for something corresponding to a new action, or a new object or attribute. Brown[8] showed that 3- to 5-year-old children were sensitive to the sentence context in picking out different meanings for *sib* across forms, such as:

Look, a sib!
Look, sibbing!
Look, some sib!

Recent work by Waxman and her colleagues[9] has shown that children also can differentiate adjectives from nouns using linguistic context:

Look, a sib!
Look, a sib one!

Naigles[10] and Gleitman[11] have pointed out how significant the sentence context is in learning a verb, in that the child can guess what kind of event is being referred to from sentence context. An intransitive verb such as:

He's sibbing

will suggest a solitary action by the agent, such as sneezing, or jumping. However, a transitive form such as:

He's sibbing him

suggests an act being done by the agent on another, such as pulling, hugging, or catching. Verbs in particular depend heavily on sentence context for the meaning to be determined because the contexts in which they occur are so rich in potentially misleading referents. Gleitman and others[10,11] emphasize that the process of learning the meaning of a new word begins with this fast mapping, a kind of informed guess about the word's meaning, but that typically several contexts are required before a meaning can be fixed.

Point 3: Words Are Organized

What is it to "fix" the meaning of a word? Crucially, it is to place it in a matrix of fine contrasts. If you have guessed that *plum* is a fruit, it is important to distinguish it from pears and apples and nectarines. If you have only a vague meaning attached to the word *carry*, for example, you don't yet have it set in contrast to slight variants such as *lift* or *hold*. Our lexicon seems to be organized into meaning clusters and contrasts, as well as elaborate networks of other associations, leading to fast and efficient retrieval. Furthermore, there is hierarchical organization to many areas, so we can call a single object by several names: a Golden Delicious, an apple, a fruit, food.[12,13] The very young child may not have such networks and contrasts so well established, and it had been noted that sometimes young children resist accepting more than one word for an object: "That's not an animal, it's a dog!"[14] However, Waxman and Hatch[15] demonstrated that children as young as 3 years old do organize their nouns into hierarchical levels. In their task, they showed children a picture and asked the children to name it at different levels of description.

For example:

Examiner shows child picture of a rose:
E: That's a dandelion.
Expected response: No it's not, it's a rose.
E: That's a tree.
Expected response: No its not, it's a flower.
E: That's an animal.
Expected response: No it's not, it's a plant.

The children were able to shift their descriptions depending on the level of the prompt. Even the 3-year-old children were sensitive to the level of the prompting in providing a contrast. So with development, the lexicon gets increasingly refined, organized into clusters, contrasts, networks, and hierarchies, making for efficient retrieval. As children reach school age, increasing demands will be placed on this organization: "Think of an animal beginning with /b/." " What's the opposite of *deep*?" "How many fruits can you tell me?"

Point 4: Semantics Goes Beyond Word Meaning

The domain of semantics is extraordinarily rich, with the meaning of individual words being just a tiny part of it. However, studies of the lexicon of young children have dominated semantic research in language acquisition until quite recently. In the last 10 or 15 years, an increasing interest has developed in the acquisition of the meaning of quantifiers.[16–19] At first glance, quantifiers (words such as *some, all, none, every, each*) may seem to pose no different a challenge than subtle contrasts in, for example, prepositions (*in, on, under, through, between*). However, quantifiers introduce the interesting problem of "scope"; that is, they have influence over other parts of the sentence. For illustration of scope take a word such as *only*:

Only the man saw a robber in a mask
The man *only* saw a robber in a mask
The man saw a robber in *only* a mask

Only "takes scope" over different parts of the sentence, and creates different meanings. Take for another example a sentence such as:

Every boy is riding a horse.

The sentence is usually understood as meaning that there are several different boys and for each boy there is a horse that he is riding. But notice it could mean that one poor horse is taking the weight of the boys all together. The sentence is ambiguous, and the ambiguity comes from whether *every* takes scope over *a*, or *a* takes scope over *every*: either,

"For every boy there is a horse to ride," or "There exists a horse, and every boy is riding it."

Research has shown that young children can readily understand these readings, but that does not mean they have the whole system mastered. Studies have shown that children entertain alternative readings that adults don't even consider,[16,18] and that they fail to limit their interpretations when readings become blocked for adults. As just one example, adults find it hard to get both readings for the sentence:

There is a horse that every boy is riding.

It seems to most adults that there must be only one beladen horse, but not to children.[20]

WHAT CAN GO WRONG

What about vocabulary learning in children with language disorders? A reduced vocabulary is a very typical sign of a language disorder, with the usual measure again being a standard picture-choice test or a naming task.[21] As a consequence, a language-disordered child will be at a disadvantage in everything from ordinary conversation to school readiness and reading. But why does the problem occur? Recent work has suggested that the problem may stem from several sources.

First, a child with a language disorder may have difficulty learning new words in a casual way, that is, from context. Research suggests that children who are language-delayed might have difficulties learning new words from such casual contexts, and may need more exposures to narrow down a word's meaning.[22] Furthermore, it is evident that a child who has grammatical difficulty (with word order or inflections, for example) will have difficulty picking up the cues to the meaning of a word that sentence contexts provide.

This observation may shed new light on a phenomenon noted by Rice and Bode,[23] who found that language-delayed children overuse "all-purpose" verbs, called GAP verbs. That is, the children studied used many more verbs of a too-general nature, such as *make, do, put, let*, and *go*, where more specific verbs might be more appropriate, such as *draw, catch, fill, help*, or *drive*. This suggests one problem may lie in retrieving words that are specific enough to meet the communicative demands. It seems important to ask whether the difficulty stems entirely from inefficiency in fast mapping meanings in the first place, or whether there is an additional difficulty in the adequacy of the child's lexical organization once the words have been learned. That is, how flexible is the retrieval of known words? Are they suitably organized into hierarchies and contrasts?

With regard to the learning of quantifiers and their scope, rather little has been done because the area is so new to the field of language acquisition. Yet an important start was made on this question in the work of Finneran.[24] With a small sample of language-delayed children, she studied the phenomena that were beginning to be explored in typically developing children. The surprising fact was that her children with language disorders were still exhibiting some of the same problems at 9 and 10 years of age. These results have since been extended in German by Penner and others.[25] Again, it is important to uncover which of the problems might stem from failure to understand the meaning of quantifiers, and which might stem from a failure to integrate that semantic knowledge with the syntax to get the right scope.

ITEM GOALS

In developing the specific probes for a linguistically fair assessment, we tapped only a tiny portion of the phenomena under the umbrella term of "semantics." In traditional tests, semantics is construed mostly as word meaning, notably vocabulary. However, specific vocabulary is precisely what a dialect-sensitive assessment seeks to avoid. We argued that finding a vocabulary test that is fair to every child regardless of culture and circumstance is an impossible task. Existing vocabulary tests are testaments to that failure.

In the Semantics domain, we consider several aspects of semantic functioning that are neglected in existing tests, but show promise for distinguishing children with language disorders from those who differ only by dialect. In trying to minimize the impact that

having different vocabularies might have on children's performance, we needed to find alternative strategies for distinguishing children who have language problems from those who do not. There are four properties of semantic functioning that we tap into in this domain—all are very different and unique to our work.

The core principles of the item types that we recommend for semantic assessment are as follows:

1. They avoid the bias of acquired vocabulary tests as too culturally dependent.
2. They look at process: Can the child learn a new word easily from context?
3. They look at lexical organization/retrieval, which may be more significant than size of vocabulary.
4. They explore in greater depth at least one element of complex semantics vocabulary: the logical properties and scope of the word *every*.

In the following text, each item will be discussed in turn.

Fast Mapping

The first subdomain to be discussed examines the process of learning a new word from context. The fast-mapping items capitalize on the essential aspect of semantic functioning that should be intact in the normal child regardless of circumstance: the ability to guess the likely meaning of a new word after just a couple of exposures. Verbs were the word type chosen, for several reasons mentioned above. First, verb learning is most heavily dependent on linguistic context for deriving a meaning.[11] Second, verb vocabulary may be less tied to variations in culture than is noun learning. Third, language-disordered children seem to have weak verb vocabularies.

The assessment that we developed was based on recent experiments that have shown that children can use syntactic information to learn something about a new verb. Even with one exposure, children are able to gain a rough meaning of a verb. Fisher[26] proposes that children can use something as vague as the number of arguments in a sentence to select between possible meanings of a verb that are carried by the transitive and intransitive frames in which the verb appears. Fisher showed 3- and 5-year-old children a video of an action. For example, one picture was of Person B sitting on a swivel stool, being spun by Person A pulling off a scarf wrapped around the waist of Person B. This action was labeled with either a transitive (a), or intransitive (b):

a. She's mooping *her* over there, or
b. She's mooping over there.

If the label for the action was (a), the action described must be the action of pulling on the scarf, and Person A must be doing the *mooping* to Person B. If the subject heard (b), the action described is most likely, though not inevitably, the action of spinning on the stool, and this time Person B is *mooping*. After three presentations, the experimenter brought out a still picture from the video and asked the child to point to the one performing the *mooping*. When given a transitive sentence (a), all of the children performed almost perfectly, selecting the agent as the subject. Children have an "agency bias," but when given an intransitive sentence (b), even 3-year-old children could override this bias to some degree, quite often selecting the patient (e.g., the one sitting on the stool) as the one doing the *mooping*.

Johnson[27] extended the basic methodology of such studies to ask whether other sentence frames would also be used to establish which action to attend to in a complex scene. She used not only transitive and intransitive verbs, but also verbs in dative constructions:

c. The boy *temmed* the flowers to the girl.

She also used complements:

d. The girl *temmed* the boy to send the flowers.

Notice that the first can only be some kind of "transfer" meaning, like *pass*, or *give*, and the second can only be some kind of communication or desire-type meaning, such as *asked* or *wanted*. The children saw three scenes in a series, as in Figure 1, in which a girl was clearly signaling to a boy with some flowers across a

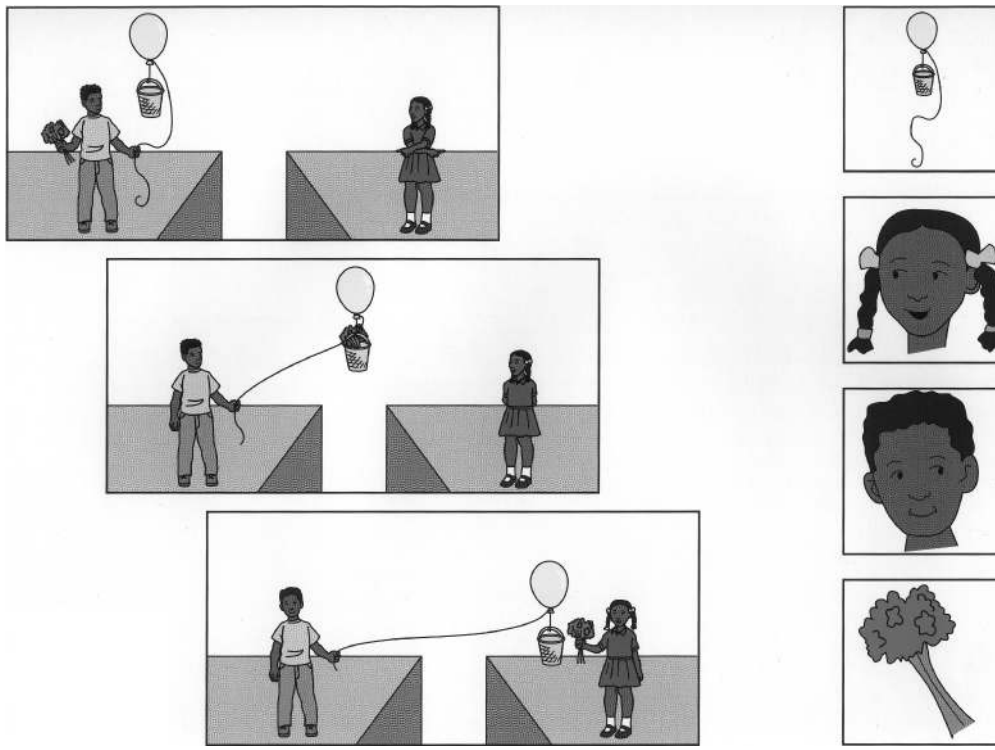


Figure 1 Fastmapping novel verbs by Dialect.

gap, then the boy floats the flowers across to her in a basket hanging from a balloon. Obviously (c) draws attention to the balloon carrying or transfer act and (d) to the signaling. The argument structure sets limits on the verb meaning. In Johnson's study, one group of children would receive one form, whereas the other group received the other form. In that way, it could be determined whether there was any inherent bias in the scenario leading a child to pick one action more often than the other as the referent. The different choices should just be governed by the argument structure.

Unlike Fisher's study, because the events were complex it was not always possible to simply point to the action being named. In the case of the complement-taking verb, the most likely meaning is a mental state or communication: You cannot easily point to *wanted* or *asked*. So, the task is designed to ask children questions about the event, and their answers reveal which action/event type they have associated with the novel verb. For example, the child is asked, "Which one is *temming*"? or "Which one got *temmed*"? However, the problem occurs if the child fails to answer the questions correctly because of the morphology or syntax of those questions, not because of the novel verb meaning? The test is introduced and modeled first with real verbs that the children know, and they are asked parallel questions about those verbs. This ensures that they have practice with the task with familiar verbs first, and also that the examiner can tell if the questions cause differential difficulty with the novel or familiar verbs.

Johnson's subjects were 60 AAE- and MAE-speaking 4- to 6-year-old children, and the task helped to reveal how much children could learn from different structures about verb meanings on limited exposure. Furthermore, the children from both dialects did equivalently well, suggesting this might be an unbiased methodology suitable for a linguistically fair test.

Johnson's methods and pictures were adapted for the items in the fast-mapping probe. The argument structures used were transitive, dative, and infinitival complement forms, each presented with a different novel verb.

Because it was not desirable to have two forms of the task, arbitrary choices were made to select one picture to associate with the transitive and another with the dative, for example. Johnson's study had shown that each action/event was equally available in the scenarios, and the child's choice was governed by the sentences used. The novel verbs are preceded by items with real verbs so the child learns the task, and so the examiner can tell if the questions cause problems. In addition, to break up response sets and provide some variety, foil questions are interspersed, such as, "Which one is wearing a green dress." More details of procedure, design, and results are provided in Johnson, de Villiers, Seymour, and Roeper (manuscript in preparation).

For time constraints, there was further selection of the materials to sample all the types of argument structures and questions in the most efficient manner, reducing redundancy while keeping the task understandable to the youngest children. As in all of the subdomains, it was necessary to balance constraints of time with concerns about breadth and diversity of tasks. In addition, it was important that some questions would tax the skills of 9-year-old children, yet not discourage the 4-year-old children because the task was too difficult. These are constraints that test-makers are familiar with but experimenters rarely encounter.

Lexical Organization

To tap lexical organization, we were inspired by the work of Waxman and Hatch[15]. However different the child's experiences and consequent vocabulary are, the child must have words organized to make efficient use of them. Words are organized in hierarchies of increasing specificity like

animal → dog → poodle
food → vegetable → lettuce

They are also organized as oppositions, such as *hot/cold*, *tall/short*, *nice/nasty*, and as synonyms or related words, such as *big/tall/large/huge*, and so forth. We knew from existing work that children begin this process of organization at about age 3 years, and their ability to relate words flexibly is a developmental process continuing into the school years. However, much of the previous research work[28] was with noun hierarchies, and nouns are perhaps the most subject to cultural variation. What if a child has not grown a garden, or been to a zoo or a dog show, or had a varied diet?

A pilot study by Pearson, Wagner, Asplin, and de Villiers,[29] asked if similar information could be gathered by using the verb domain instead of nouns. Verbs are a more culturally neutral domain to examine because the majority of common verbs do not require specific cultural experiences. Children raised in poverty, therefore, may have relatively more equal opportunities to acquire a structured vocabulary of verbs than they do of nouns.[4,30] Furthermore, there is suggestive evidence that children with language impairments may have reduced verb vocabularies, with an increased use of all-purpose verbs.[23] Thus verbs are a promising domain in which to differentiate between language disorder and difference.

Verbs show less hierarchical organization than do nouns but it is still present in some sets:

move → walk → stagger

That is, walking is a type of moving; staggering is a specific manner of walking. In the verb contrasts subdomain we ask whether a child can provide an appropriate contrast, at the appropriate "level" in the hierarchy, in naming some actions in flexible ways. We created a verb antonym task that was modeled after the Waxman and Hatch[15] study of noun hierarchies. Linguistic research has suggested that verb hierarchies are relatively "flat."[31] Nevertheless, an antonym task with verbs does seem to tap lexical subcategories, such as manner of motion, direction of motion, means of creation, etc.[32,33]

For example, to a picture of a girl licking a pop sicle, we might say

"This girl is not chewing the ice cream, she's. . ."

an appropriate response is "licking," but not "eating," even though that is also true of the picture. However, given the prompt:

"She's not drinking the ice cream, she's. . ."

then "eating" is a better response than "licking." To a picture of a man hammering a toy airplane together, we would say:

"He's not gluing the airplane, he's. . ."

and expect, for example, "hammering it," not "fixing it." But if we then said,

"He's not breaking the airplane, he's . . ."

we expect, for example, "fixing it," not "hammering it." To succeed at this task, children need to have some minimum number of verbs in their vocabulary, and, equally importantly, they must have those verbs organized into appropriate subcategories and contrasts.

The task consisted of 10 pictures of a range of common verb classes, such as motion, grooming, and breaking, that have at least partial hierarchies.[33] A single practice item was used to demonstrate the task above, with three different prompts per picture: (a), (b), and (c), differing in what they picked out. Often, but not always, the prompts increased in generality from (a) to (c). If the child failed to produce a new verb to a further prompt, the experimenter said, "You already said that—think of something else," and repeated the prompt. Only one such repeated prompt was given at each step.

We coded the adequacy of the child's answer; namely, was it at the right level of description? We also coded the novelty of the answer: Did the child just repeat the same answer or give three different answers? In this work we attempted a refined scale, with 0 being inadequate; 1 being on target, but too specific or general; and 2 being just right.

SAMPLE TOO-SPECIFIC ANSWERS

In answer to:

"He's not breaking the car apart, he's. . ."

A child might say, "painting the door."

"She's not hating the pop sicle, she's. . ."

A child might say, "licking it."

SAMPLE TOO-GENERAL ANSWERS

In answer to:

"He's not gluing the airplane, he's. . ."

A child might say, "holding it,"

"He's not breaking the car apart, he's. . ."

A child might say, "working".

A score of 2 was given for a good antonym of the verb provided in the prompt, such as *licking* not *chewing*, or *painting* not *gluing*.

The preliminary work with 30 AAE-speaking children aged 4 to 6 years suggested the feasibility of the task and measures, and that there was development throughout that age range. A very small pilot study with four language-disordered children suggested the task might discriminate well between normal and disordered children.

The pictures and prompts were adapted for the experimental test, with a larger sample of items initially that was reduced for the reasons of time and efficiency. Instead of three prompts per picture, which sometimes strained even our verb lexicons, pictures were chosen that were suitable for two clear prompts. Prompts needed to have distinct answers, because in a production task it is often surprising how much variability there is and how much ambiguity of scoring can then result (especially over this very large sample and age range). The fine gradations were abandoned for consistency and ease of scoring, so the new scoring is simply 1 or 0 for each item depending on its adequacy as a contrast.

The probe was successful enough that we sought to use the same task design to look at another domain of semantic contrasts, namely prepositions. Prepositions are also in a matrix of oppositions, though again without much hierarchy. A child must know the subtle differences between prepositions to arrive at the right description of a scene. Work on prepositions had revealed that although prepositions are used early, preposition contrasts emerge over the course of development.[34,35] Furthermore, even closely related European languages set up

the contrasts among spatial prepositions in different ways,[34] so the child's task is not so easy. In addition, not all prepositions refer to space: some are abstract or purely grammatical in function. For example, why do we listen *to* something, look *at* something, go *on* vacation? The challenge for the child is to learn how the language uses prepositions in both concrete and abstract ways.

The preposition task mirrors the verb contrast task in providing prompts for the child, two per picture, that require the child to fix or supply a preposition to describe what is happening. For example, the child is shown a picture of a girl riding on a horse.

The prompts would be:

"She's not riding to a horse, she's riding. . ."

The expected response would be "*on a horse*"

"She's not sitting *behind the saddle*. . ."

The expected response would be "in the front seat"

The adequacy of the child's responses is fairly easy to judge. We avoided cases where we had reason to believe AAE might introduce a difference, though we may not have succeeded entirely because so little is known about dialect differences in children's use of prepositions. Orr[36] reports a use of *at* in AAE in place of Mainstream American English (MAE) *to* that we may not have been sufficiently sensitive to in the current scoring; for example, "She's going at Grandma's."

The Meaning and Scope of the Word *Every*

We argued above that semantics is not just about the lexicon, it also involves the way that quantification works in sentences. Quantifiers are terms like *less, more, some, any, all, every, each, none*, and the way these terms interact in sentences to give meaning is quite intricate.

The work leading up to the design of the items in these subdomains is quite new and its interpretation is still controversial. For example, Philip[17] and Drozd[18] have both found that children make systematic misinterpretations of a sentence such as

"Is every man on a horse?"

If the scene contains an extra horse, children say "No, not that one," pointing to the horse. It has been suggested[17,20] and contested[19] that children treat "*every*" as if it means "always" (very roughly speaking); that is,

"Is it true in every case that a man is on a horse?"

The tantalizing idea is that *every* at first is not anchored to the noun phrase "man," but somehow takes scope over the whole sentence. In addition, previous work had shown that children did not restrict the scope of "*every*" to elements within its sentence, but sometimes allowed it to cross sentence boundaries. Take for example, the following sentences:

The man watched *every* cat. *He* scratched his ear.

For an adult, *he* cannot possibly refer to *every cat*. This is not because the pronoun is singular, because if a singular form (e.g., *his*, or *a*) occurred within the same sentence as *every*, it would be fine to give it that multiple or "variable" interpretation:

The man watched *every* cat scratch his ear.

So the child must learn the meaning of specific quantifiers; the differences between *all*, *some*, *none*, and *every*. The child must recognize that these modify nouns, not whole events; for example, *every* is not the same as *always*. The child must also learn the limits on how *every* works within a sentence, not across sentences:

The man watched every boy play the drum,

is not the same as:

The man watched every boy. He played the drum.

Finneran[24] had tested a variety of different quantifier and *wh*-question tasks that involve these subtle semantics on a small group of

MAE-speaking language-disordered children, and had evidence for significant delay in their understanding even at age 9 and 10 years. However, we had no guarantee that AAE development might work the same way as MAE, and several reasons to suspect it might not. AAE has properties such as negative concord, which can be construed as a form of quantification that takes a very different form syntactically. We therefore included several different tests of the quantifier *every* to determine if the tests could be both fair across dialects and discriminating of difference.

The quantifier items are restricted to evaluating children's understanding of the word *every* and how it works in sentences. In the simplest of cases, the child is asked about a picture in which all but one man, for example, would be driving a car:

Is every man driving a car?

The answer is "no" if the child understands that *every* does not mean, for example, *some* or *any*.

More interesting is a picture in which several monkeys are eating bananas, and a rabbit is eating a carrot. The question is:

Is every monkey eating a banana?

The answer is "yes," if the child understands that *every* is only attached to *monkey*, and not to other cases like the rabbit. If the child thinks that the sentence means that every subevent had to be one of a monkey eating a banana, the child would answer "no."

Finally, the child sees a pair of contrasting pictures that give two alternative readings for sentence sets such as:

The man watched every boy. He played the drum.

The child must choose the picture in which the man, not every boy, is playing the drum. If the child chooses the correct picture, he or she understands that *every* cannot cross sentence boundaries for *he*.

Children can fail in several ways on the quantifier tasks. For example:

1. They can miss the basic meaning of *every* and answer "yes" instead of "no" to the car picture.
2. They can take *every* to apply also to the rabbit in the monkey picture, not limiting it to the noun phrase *monkey*, and thus say "no" instead of "yes."
3. They can fail to discern the condition on the scope of *every* imposed by a sentence boundary, and think *he* refers to every boy in the case of "The man watched every boy. He played the drum."

HOW CHILDREN PERFORM ON THE SEMANTICS SUBDOMAINS

How Children Perform on Fast Mapping

The fast-mapping task produces intriguing data at the level of the individual child, question, and construction type that are too elaborate to describe here; the reader is referred to Johnson et al. (manuscript in preparation) for the full details. Instead, what follows is a summary of the composite data across the different age, dialect, and disorder-status groups. The child is given a point for a correct answer judging by the adult grammar, and the points are tallied across constructions and questions.

Figure 2 shows that the data from this study replicate Johnson's initial work. Even with this much larger sample, and larger age range (4 to 9 instead of 4 to 6 years), there is no measurable difference in the performance of AAE- and MAE-speaking children. [Novel Verbs: Age, $F(5, 1002) = 15.141$, $p < 0.0001$; Dialect, $F(1, 1002) = 0.415$, $p = 0.520$; Age by Dialect, $F(5, 1002) = 1.150$, $p = 0.332$.]

Furthermore, there is a very reliable age trend that continues throughout the age range, making this a suitable item-type for examining development. However, Johnson's methods and procedures were selected primarily because they suggested (and now we confirm) the lack of bias in the test. If the test shows no reliable difference between normally developing and disordered subjects, it will fail the other requirement of our test.

Fortunately, Figure 3 shows that the test is much more challenging for language-disordered children at every age. [Novel Verbs:

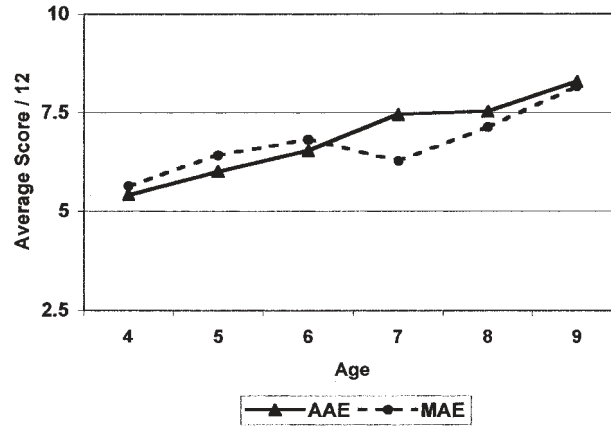


Figure 2 Quantifier *every* items by Dialect.

Age, $F(5, 1002) = 16.328$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 42.567$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 2.146$, $p = 0.06$.] In our fuller account of these data, we discuss where the differences are located (Johnson et al, manuscript in preparation).

How Children Perform on Verb Contrasts

The verb contrast task lived up to the expectations generated by the pilot study, over the larger age range. Normally developing children who spoke AAE or MAE proved increasingly adept with age at providing an appropriate level contrast. In addition, there was no significant difference in the performance of the two dialect groups. (See Figure 4) (Age, $F(5, 1002) = 57.116$, $p < 0.0001$; Dialect, $F(1,1002) = 6.885$, $p = 0.009$; Age by Dialect, $F(5, 1002) = 0.838$, $p = 0.523$.) Note that the main effect of Dialect is almost entirely due to the 4-year-old children, and is nonsignificant once they are removed [$F(1, 796) = 3.576$, $p = 0.06$.] Mistakes, especially for the younger children, consisted of saying things such as, "I don't know," but more commonly giving an answer that was too specific or too general to be the right level for the prompt. This could suggest that young children's verb vocabularies either are still incomplete or weakly organized.

However, it is very clear from Figure 5 that the problems that young children have are amplified in language-disordered children, who differ from their normal age-mates at all ages. In analyses of variance: Age, $F(5, 1002) = 65.505$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 85.616$, $p < 0.0001$; Age by Clinical

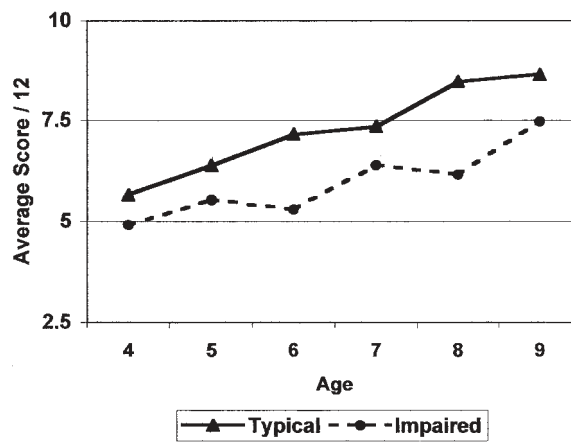


Figure 3 Fast mapping novel verbs by Clinical Status.

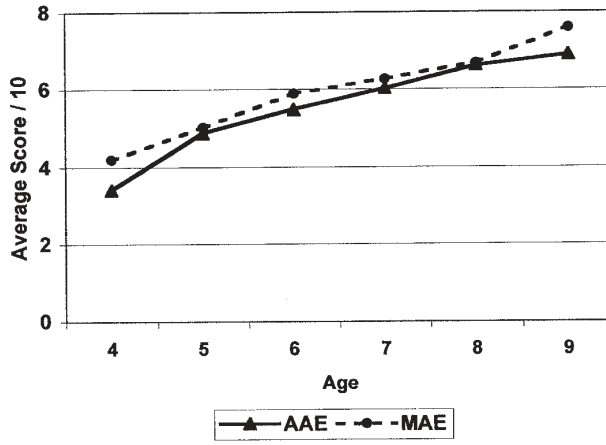


Figure 4 Verb contrasts by Dialect.

Status, $F(5, 1002) = 0.524$, $p =$ not significant. The disordered children use more vague expressions, such as all-purpose verbs (*do*, *try*, and *put*) or they omit the verb. They have difficulty with the right contrasts after the prompt, and have difficulty saying two different things about the same picture.

Thus, the subdomain discriminates well between children with language disorder and children who are normally developing, and confirms previous reports that the verb vocabulary may be vulnerable in language-disordered children.

How Children Perform on Preposition Contrasts

The preposition contrast task did not have the same level of pilot work behind it to back up the hints in the literature that it might prove revealing. Nonetheless, the data from the field testing (Fig. 6) showed good age trends over ages 4 to 9 years, and equivalent levels of performance from MAE- and AAE-speaking children [Age, $F(5, 1002) = 18.379$, $p < 0.0001$; Dialect, $F(1, 1002) = 1.077$, $p = 0.30$; Age by Dialect, $F(5, 1002) = 0.682$, $p = 0.637$]. Mistakes consisted of confusing prepositions or sometimes omitting them altogether. These mistakes, especially omissions, were much more prevalent in the disordered children, who can be seen clearly to lag behind their normally developing age-mates in Figure 7. [Age, $F(5, 1002) = 20.187$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 35.288$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 0.914$, $p = 0.471$.] The disordered children sometimes use odd forms, such as, "down the chair" instead

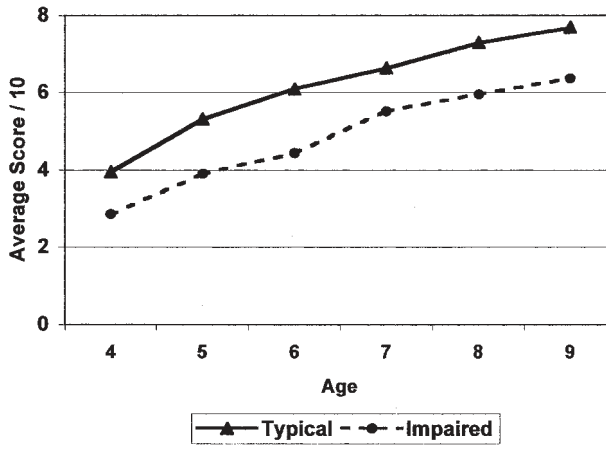


Figure 5 Verb contrasts by Clinical Status.

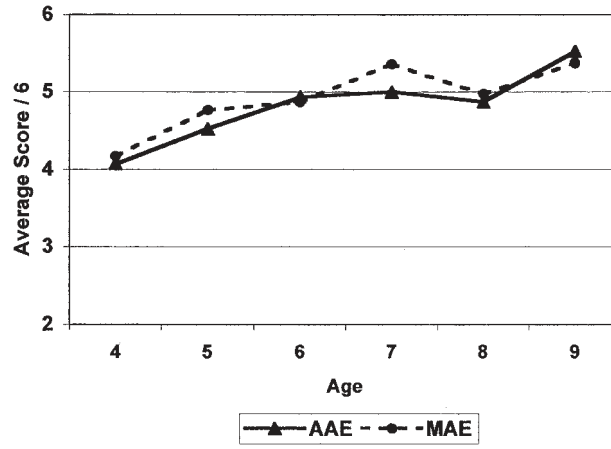


Figure 6 Preposition contrasts by Dialect.

of "under the chair," or "up the cat" instead of "up to the cat." We have not detected any particular difference between spatial and abstract prepositions, but the sample of the latter is tiny.

How Children Perform on the Quantifier Subdomain

First, consider the items about the meaning of *every* within a sentence. The results of the field testing research reveal that the findings of earlier studies are borne out. Children do make the kinds of mistakes we have highlighted above, and there is a clear developmental time course persisting until age 7 or 8 years even in normally developing children (Fig. 8). We still need to explore these data in detail, but for the moment, these items appear to fulfill our requirement not to create bias against AAE speakers. In the analyses of variance, the values for these items followed the general pattern of the other item-types: Age, $F(5, 1002) = 33.051$, $p < 0.0001$; Dialect, $F(1, 1002) = 2.225$, $p = 0.136$; Age by Dialect, $F(5, 1002) = 0.877$, $p = 0.496$.

In addition, the subdomain is highly successful in separating out disordered children from either dialect group. It is clear that the language-disordered children are considerably delayed in mastering this quantifier (Fig. 9) (Age, $F(5, 1002) = 31.766$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 27.675$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 0.351$, $p = 0.882$).

With regard to the sentence boundary items, Figure 10 reveals a good match in the developmental time course for AAE and MAE children [Age, $F(5, 1002) = 44.981$, $p < 0.0001$;

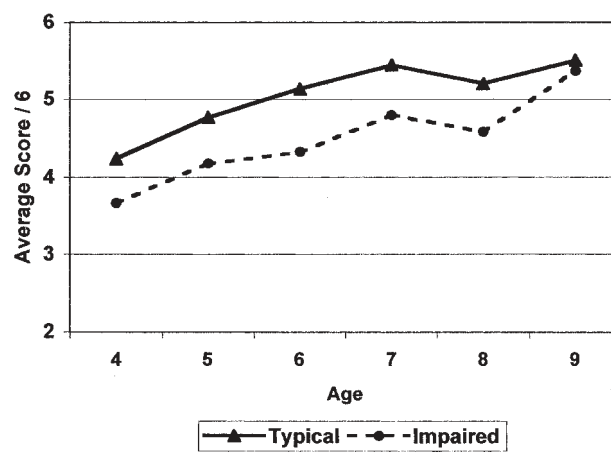


Figure 7 Preposition contrasts by Clinical Status.

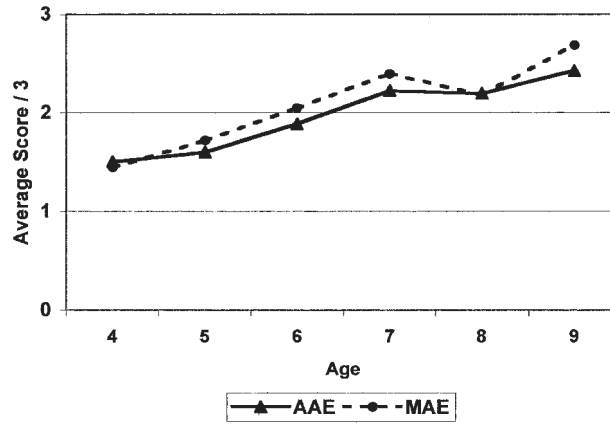


Figure 8 Quantifer *every* items by Dialect.

Dialect, $F(1, 1002) = 0.051$, $p = 0.821$; Age by Dialect, $F(5, 1002) = 0.946$, $p = 0.450$].

Figure 11 shows the expected lag for language-delayed children, at least after age 4 years [Age, $F(5, 1002) = 42.180$, $p < 0.0001$; Clinical Status, $F(1, 1002) = 34.711$, $p < 0.0001$; Age by Clinical Status, $F(5, 1002) = 3.408$, $p = 0.005$].

HOW DO THESE FACTS HELP US IN FULL DIAGNOSIS AND REMEDIATION?

To an experimenter, test making is a peculiar process. Experiments demand many examples to test the robustness of an effect; tests require the minimum that will provide group separation, for efficiency and time. Experiments demand keeping separate those types that are significant theoretically; tests lump items together to give general descriptions. Ironically, however, once the testing is done and the child has a "diagnosis" of a language disorder, the same questions should occur to the therapist as occur to the experimenter. Just why did the child fail this item? What does this pattern of responses mean? How do I intervene and fix the problem if I only know there is a general one?

Thus, although a test requires giving points for different sections and adding them up, we think it is important to examine the details of what the child did on each item-type to determine the source of the problem.

Fast Mapping

Consider fast mapping. When the child fails, it is important to know several things. Did the

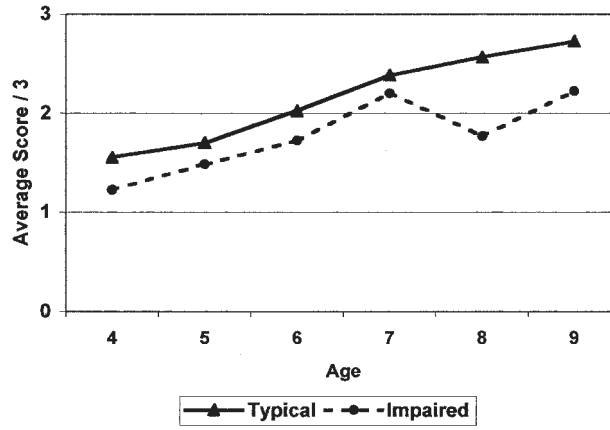


Figure 9 Quantifier *every* items by Clinical Status.

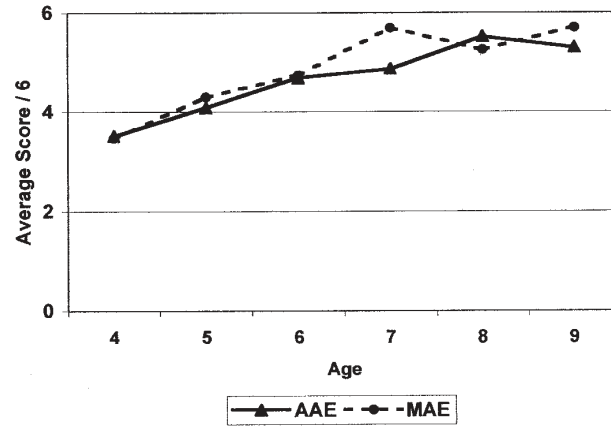


Figure 10 Quantifier sentence boundary scores by Dialect.

child fail on the real verbs as well as the novel verbs? If so, what kinds of questions were difficult? A substantial failure on the real verb questions is more suggestive of a problem with syntax or morphology of those questions than it is of a failure to learn a novel verb. Suppose the child passes most of the real verb questions, but fails on the novel verbs. Does the child tend to fail on the more complex argument structures, such as datives and complements? Again, the suggestion is that the child may have trouble learning new verbs if the sentence structure in which they appear is too complex. How is the child's performance on the syntax item-types? Perhaps the child has trouble across the board with novel verb learning. Is the presence of a novel verb affecting his or her memory? How does he or she perform on nonword repetition in the phonology domain of the Screener, Part 2?

Verb Contrasts

Next consider verb contrasts. How does the child's performance on verb contrasts compare with his or her performance on fast mapping? Is the child showing evidence of being a slow learner from context? Is the failure a genuine problem with the size of his or her verb vocabulary? What kinds of errors occur? If the verbs used are too general, perhaps the child has a limited verb vocabulary. If they are too specific, then the child may have a disorganized lexicon. If the child "misses the mark" of the pictures, does this show up in the preposition task or on the pragmatics tests, where that skill

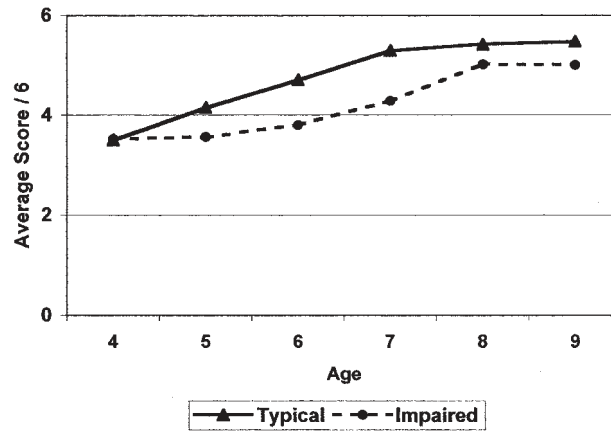


Figure 11 Quantifier sentence boundary scores by Clinical Status.

of getting the meaning from a picture is at a premium?

Preposition Contrasts

In the preposition task if the child omits prepositions, how is his or her morphosyntax on the screener? How is his or her memory for phonological material in phonology nonword repetition? Is there evidence from the article task (syntax) that the child has a general difficulty with functors, that is, small grammatical words? If the child provides odd prepositions, does this parallel a problem in verb contrasts?

Quantifiers

In the quantifier subdomain, does the child tend to say "yes" to everything? Does the child show a basic understanding of the meaning of *every* on the car picture? If not, perhaps the basic meaning is still absent, and work needs to be done to contrast *all, some, every*, and *one*. Does the child insist that the rabbit is relevant in the monkey/banana picture? If so, perhaps the meaning of *every* is not yet linked to its noun phrase. How do such children perform on the double *wh*-questions in *wh*-syntax and *wh*-production? Might the child have a problem with the notion of scope or variables? In the sentence boundary items, does the child show a position bias in choosing pictures? That would be the most primitive error. If not, does the child get the single sentences right? If so, does the child take *every* to cross the sentence boundary? How does the child's problem relate to his or her understanding of other complex syntax; for example, passives or *wh*-questions?

CONCLUSION

This article has demonstrated some ways it might be possible to develop relatively unbiased methods for assessment of the process of vocabulary learning and the efficient organization of vocabulary as alternatives to the measurement of the size of acquired vocabulary. It has also shown that semantic knowledge extends beyond word meaning conceived simply in terms of content words, into areas such as quantifier scope. This latter knowledge may be especially sensitive to language disorders, and provide a window into larger aspects of the grammar as a whole. There is still some distance to go in translating the discoveries into careful diagnostics that will allow efficient and precise remediation, but the prospect is an exciting one.

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. Stockman I. The new Peabody Picture Vocabulary Test-III: an illusion of unbiased assessment? Lang Speech Hear Serv Schools 2000;31:340–353

2. Allen B, Boykin AW. African-American children and the educational process: alleviating cultural discontinuity through prescriptive pedagogy. School Psych Rev 2000;29:586–593
3. Boykin AW, Allen B. Beyond deficit and difference: psychological integrity in developmental research. In: Yeakey CC, ed. Edmund Gordon: Producing Knowledge, Pursuing Understanding. Stamford, CT: JAI Press; 2000:15–34
4. Blake I. Language Development in Working-Class Black Children: An Examination of Form, Content, and Use [dissertation]. Columbia University, New York; 1984
5. Choi S, Gopnik A. Early acquisition of verbs in Korean: a cross-linguistic study. J Child Lang 1995;22:497–529
6. American Speech-Language-Hearing Association. Position paper: social dialects and implications of the position on social dialects. ASHA 1983;25: 23–27
7. American Speech-Language-Hearing Association Technical report: American English Dialects. ASHA. In press
8. Brown R. Words and Things. Glencoe, IL: The Free Press; 1958
9. Klibanoff RS, Waxman SR. Basic level object categories support the acquisition of novel adjectives: evidence from preschool-aged children. Child Dev 2000;71:649–659
10. Naigles L. Children use syntax to learn verb meanings. J Child Lang 1990;17:357–374
11. Gleitman L. The structural sources of verb meanings. Lang Acquis 1990;1:1–27
12. Rosch E, Mervis C. Family resemblance: Studies in the internal structure of categories. Cognit Psychol 1975;7:573–605
13. Mervis C, Crisafi M. Order of acquisition of subordinate-, basic-, and superordinate-level categories. Child Dev 1982;53:258–266
14. Inhelder B, Piaget J. The Early Growth of Logic in the Child. New York: Norton; 1964
15. Waxman SR, Hatch T. Beyond the basics: preschool children label objects flexibly at multiple hierarchical levels. J Child Lang 1992;19:153–166
16. Philip W. Event Quantification in the Acquisition of Universal Quantification [dissertation]. University of Massachusetts, Amherst, MA; 1995
17. Philip W. The event quantificational account of symmetrical interpretation and a denial of implausible infelicity. In: Stringfellow A, Cahana-Amitay D, Hughes E, Zukowski A, eds.Proceedings of the 20th Annual Boston University Conference on Language Development, Somerville, MA: Cascadilla Press; 1996:564–575
18. Drozd K, van Loosbroek E. Weak quantification, plausible dissent, and the development of children's pragmatic competence. In: Greenhill A, Littlefield H, Tano C, eds. Proceedings of the 23rd Annual Boston University Conference on Language DevelopmentSomerville, MA: Cascadilla Press; 1999: 184–195
19. Crain S, Thornton R, Boster C, Conway L, Lillo-Martin D, Woodams E. Quantification without qualification. Lang Acquis 1996;5:83–153
20. Roeper T, de Villiers JG. The emergence of bound variable structures. In: Reuland E, Abraham W, eds. Knowledge and Language: Orwell's Problem and Plato's Problem. Boston: Kluwer Academic; 1993:105–139
21. Leonard L. Specific language impairment: characterizing the deficit. In: Levy Y, Schaeffer J. Language Competence across Populations: Toward a Definition of Specific Language Impairment. Mahwah NJ: Lawrence Erlbaum; 2003:209–232
22. Rice M. Synthesis and commentary. In: Rice M, Schielfelbusch R, eds. The Teachability of Language. Baltimore, MD: Paul Brookes; 1989:351–355
23. Rice ML, Bode J. GAPS in the verb lexicons of children with specific language impairment. First Lang 1993;13:113–131
24. Finneran DA. Bound Variable Knowledge in Language Disordered Children [thesis]. University of Masschusetts, Amherst, MA; 1993
25. Roeper T, Pearson BZ, Penner Z, Schultz P. The Emergence of Wh-Variables: Cross-linguistic Explorations. Paper presented at: IX International Congress for the Study of Child Language (IASCL) and the Symposium on Research in Child Language Disorders (SRCLD), University of Wisconsin, Madison, WI; July 16–21, 2002
26. Fisher C. Structural limits on verb mapping: the role of analogy in children's interpretations of sentences. Cognit Psycol 1996;31:41–81
27. Johnson V. Fast mapping Verb Meaning from Argument Structure [dissertation]. University of Massachusetts, Amherst, MA; 2001
28. Anglin J. The Growth of Word Meaning (Research Monograph No. 63). Cambridge, MA: The MIT Press; 1970
29. Pearson BZ, Asplin K, Wagner L, de Villiers J. Organization is the key: alternative measures of the verb lexicon. Poster session presented at: Symposium on Research in Child Language Disorders, Madison, WI; June 7–9, 2001
30. Stockman I, Vaughn-Cooke FB. Addressing new questions about black children's language. In: Fasold R, Schiffrin D, eds. Current Issues in Linguistic Theory: Language Change and Variation. Vol. 52. Philadelphia: John Benjamins; 1989:275–300
31. Miller G, Fellbaum C. Semantic networks of English. Cognition 1991;41:197–229

32. MacWhirter S. Subcategories of Action Verbs in Children's Language [thesis]. Michigan State University, East Lansing, MI; 1989
33. Levin B. English Verb Classes and Alternations: A Preliminary Investigation. Chicago: University of Chicago Press; 1993
34. Bowerman M. Learning how to structure space for language: a cross-linguistic perspective. In: Bloom P, Peterson M, Nadel L, Garrett M, eds. Language and Space. Cambridge, MA: MIT Press; 1996: 385–436
35. Johnston JR, Slobin D. The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. J Child Lang 1979;6:529–545
36. Orr E. Twice As Less: Black English and the Performance of Black Students in Mathematics and Science. New York: Norton; 1987

A Noncontrastive Model for Assessment of Phonology

Harry N. Seymour, Ph.D.[1]

ABSTRACT

The contrastive/noncontrastive model developed by Seymour and Seymour (1977) can be applied to the development of a dialect-sensitive phonological assessment that uses a single scoring and test format, regardless of a child's dialect. Through extensive field research, stimulus items were found that respect the phonotactics of African American English (AAE) (i.e., no targets are final consonants or final consonant clusters), yet are sufficiently demanding to show development in the age range from 4 to 9 years. Consonant clusters of varying levels of phonological difficulty are shown to discriminate between typically developing and phonologically impaired children of different dialect groups, Mainstream American English (MAE) and non-MAE. Implications for further diagnosis and remediation are presented.

KEYWORDS: Contrastive and noncontrastive substitution patterns, type I, II, and III errors, consonant clusters

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) how the contrastive-noncontrastive model of assessment was applied in the development of the proposed dialect-sensitive phonology assessment, and (2) several key features of the proposed phonology assessment items.

PHONOLOGY DOMAIN

The contrastive/noncontrastive model[1] described in Seymour[2] was applied in the development of a dialect-sensitive phonology assessment. As with other domains of language, *contrastive* refers to phonological features that differ (contrast) between African American English (AAE) and Mainstream American English (MAE). The most common of these features in AAE are absent elements of consonant clusters and the absence of a word final consonant.[3,4] These contrastive features have

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: Harry N. Seymour, Ph.D., Department of Communication Disorders, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: hseymour@comdis.umass.edu. [1]Professor Emeritus, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01, 091,100,ftx,en;ssl00184x.

particular diagnostic implications. Consider the cluster "st" as an example. In AAE phonology, the reduction of /st/ → [s] is commonly observed. Although the cluster is similarly absent in the speech of very young MAE-speaking children, it is no longer absent by 4 years of age in the initial word position, and in the final word position, "st" is typically mastered by 7 years of age.[5] These age-appropriate expectations can indicate a possible phonological/articulation disorder if a child fails to meet them.

However, in the case of an AAE-speaking child who may produce either "fast" or "fas," that child should not necessarily be viewed in the same way. When spoken by children of AAE backgrounds, this type of consonant reduction should not be considered a potential clinical problem, but should be viewed as a contrastive feature of AAE. Hence, such AAE features have at best limited diagnostic value as indicators of a phonological disorder when they are absent either as single consonants in final word positions or elements of a cluster. Because of the ambiguity of their status, it would not be clear whether the absent forms are due to the AAE dialect or to a phonological problem.

Although most contrastive features produce ambiguous interpretations in the diagnostic process, this is not true of all contrastive features. Clearly, a consonant cluster reduction such as /st/ → [t] (as in "ret" for "rest") could not be explained by AAE phonological rules, and thus, would be considered either an error of development or of disorder. In addition, the ambiguity associated with absent forms does not exist for all substitution patterns. Phonological substitution patterns that do not follow the AAE characteristic profile would be patterns of either development or disorder. For example, an /f/ → [th] substitution would yield a "mouth" → [mouf] substitution in AAE, but a /th/ → [s] substitution as in "mouth" → [mous] would be inconsistent with AAE and could only be explained as a developmental or a disordered error.

As for most substitution patterns, there is no ambiguity with noncontrastive patterns. The "st" in the initial position (*st*eam) is noncontrastive and is expected to be mastered by 4 years of age. Failure to achieve this mastery would arouse concern. Although this concern would most probably apply for AAE speakers as well, there remains some question about whether AAE and MAE follow the same acquisition schedule. As stated above, age 4 is the expected MAE mastery milestone for the initial "st." Is it the same for AAE? The answer is probably "yes" based on most studies comparing MAE and AAE among children,[6,7] but these studies are limited[8]; in the absence of extensive longitudinal studies on AAE comparable to those on MAE, this question cannot be answered with absolute certainty. For this reason, we chose stimulus items based on extensive field-testing and piloting to ensure that items were in fact noncontrastive and that they would not differ between AAE and MAE in terms of age appropriateness.

RESEARCH BACKGROUND

The phonology stimulus items on which we have concentrated derived from several years of research on AAE phonology. As a result of a National Institutes of Health (NIH) grant to Seymour and Roeper in 1994 (See Acknowledgments), an extensive database on AAE phonology was developed. This database made possible an in-depth analysis of phonological constraints governing features of AAE. The phonological contexts under which AAE patterns were favored confirmed important similarities between child AAE and the adult AAE phonology described in the literature.[6,8–11] This led to the selection of more than 200 consonant and cluster combinations about which predictions were made as to which best discriminated AAE and MAE by age, which were produced in identical ways by MAE and AAE speakers across ages, and which best discriminated typically developing children from phonologically impaired children.

These predictions about potential stimulus items for phonology assessment were tested in a 1998 NIH contract awarded for 6 years to Seymour, Roeper, and de Villiers. (See Acknowledgments). The NIH contract addressed several aspects of language and involved piloting of test items in schools and preschools in Hartford, Connecticut, and Springfield, Massachusetts. Overall, several hundred

African American and non-Hispanic white children who were determined either to be AAE speakers or MAE speakers, and who were matched for socioeconomic status and educational background, participated in pilot studies. See Pearson[12] for details regarding field testing and developmental milestone research, and Ciolli and Seymour[13] for uses of the contrastive features that we piloted for screening items.

THE ASSESSMENT FORMAT

The assessment format of the Phonology domain that was finally constructed allowed two important challenges to be met. First, to ensure a developmental demarcation across ages, consonant clusters of varying levels of phonological difficulty were used as targets in the stimulus words. Second, using clusters only in the initial and medial word positions avoided AAE patterns and thus eliminated difficult clinical decisions about what is disordered and what is not among speakers of AAE. By restricting target stimuli in this way, the testing format allowed for the most difficult phoneme combinations and contained no AAE patterns. It also created a natural developmental growth curve.

Twenty-five word targets containing 25 clusters represented the target stimuli. These particular targets were selected following extensive field testing of numerous target words and clusters. They were among the best at discriminating between typically developing and phonologically impaired children. Their effectiveness stems from the following several unique phonotactic and context conditions:

1. Clusters were assessed within continuous sentence context as opposed to in isolated words;
2. Clusters were assessed in either a monosyllabic or multisyllabic context;
3. Clusters comprised both two and three consonants;
4. Clusters were either intra- or intersyllabic;
5. Some clusters were assessed within iambic words; and
6. Some clusters comprised liquids (/r/ and /l/).

The administration of the Phonology items involved the presentation of a picture depicting a cartoon image about which the child was asked to repeat a sentence produced by the examiner. Each sentence began with the carrier phrase "I see a . . ." introducing a small clause or noun phrase. For example, the examiner would show a picture of a truck and say, "I see a truck pull a boat," or a picture of a leaf and say, "I see a leaf." The child was to repeat the entire sentence exactly as presented. This procedure was an attempt to control the memory load across stimulus sentences, whose length could vary greatly. In addition, in an effort to avoid dialect interference among AAE speakers, this first-person present tense construction ("I see. . .") restricted the use of certain morphosyntactic markers. To illustrate a context in which a dialect interference is possible, consider the target sentence "The boy *cl*oses the door" containing the target cluster /kl/. The third-person singular form ("close*s*") could interfere with an AAE speaker's production of /kl/ because of the natural inclination to reduce "closes" to "close" in this third-person context. By using the small clause construction this potential interference is avoided, as in the sentence "I see the boy close the door." See Figure 1 for examples of target stimuli.

ASSESSMENT FORMAT RATIONALE

Background research for the selection of targets and the stimulus format is discussed below.

Age Difference and Diagnostic Relevance

Considerable research supports the use of consonant clusters to test phonology in this age range of 4 to 9 years. In general, children do not fully master the production of all of the consonants of English until around 8 years of age.[14–16] Even though children have acquired their phonemic system of contrasts much earlier, during this period of acquisition, there is a gradual but progressive mastery of phonetic constraints.[17] The error patterns produced by children during the acquisition period are in fact systematic and predictable.[18,19] It is when children no longer produce these developmental

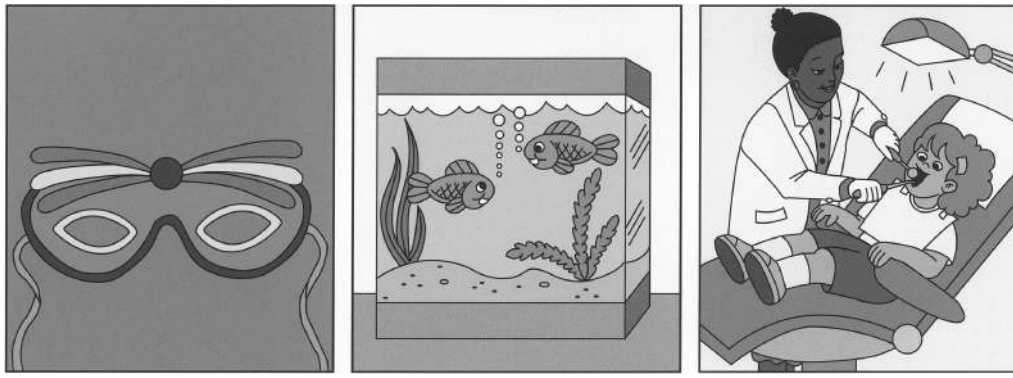


Figure 1 Phonology stimuli.

error patterns that they have achieved mastery of their phonemic system. This acquisitional mastery occurs over several years and its order is also systematic and predictable.[5] It is this predictable order that determines the age-appropriate milestone targets characteristic of most tests of phonology. Our proposed test also draws upon the predictable and systematic nature of phoneme acquisition, but is unique in that stimulus items were designed to be difficult to sharpen distinctions across ages and enhance the differences observable between typically developing and phonologically impaired children.

There is supporting evidence for each facet of the testing format: that generally clusters are more difficult to master than singleton phonological structures[20]; that three member clusters are more difficult to master than two member clusters (e.g., "*str*eam" versus "st*ick*"[17]; that intrasyllabic clusters [re – (str)ain] are more difficult to produce than inter-syllabic clusters [ce(*n*—*tr*)al][21]; that multisyllabic words are more difficult to produce than monosyllabic words[22]; that children with disordered or delayed phonology may have difficultly producing the first syllable in iambic words [(a)sleep with initial unstressed syllable];[23–25] that consonant clusters containing liquids can often result in consonant cluster simplification among speakers with immature and disordered phonological systems[17]; and that productions of words targeted for articulation are more authentic in continuous speech than in isolated words.[26–30] All of these characteristics converge in creating a relatively difficult production task that challenges less mature and impaired phonologies while maintaining strong milestone distinctions across ages.

English Variation Status

We sought a phonological assessment method for the proposed test that would work as well for AAE- as it does for MAE-speaking children. Our final design included dialect-neutral stimulus items that emphasize the identification of phonological impairment without penalizing a child for phonological patterns typical of variations in English such as AAE. Indeed, AAE phonological features were quite prevalent among most of the African American children examined in the various pilot studies and standardization sampling that we have conducted, in keeping with the extant literature. These included the most often observed patterns (i.e., absent final consonants and absent consonants of word final clusters).[3,4,6,8,21] Because no stimulus items involving consonants, final consonants, or final consonant clusters as targets were included in the final phonology probes, AAE features were inconsequential because there is no requirement for a child to produce them under test conditions. Thus, the diagnostic dilemma faced by most phonology tests is avoided (i.e., having to adjust and manipulate scores for AAE speakers compared to MAE speakers). This kind of manipulation attempts to accommodate different response patterns for AAE and MAE, creating the complication of a possible different scoring format and/or scale for each of those variations of English.

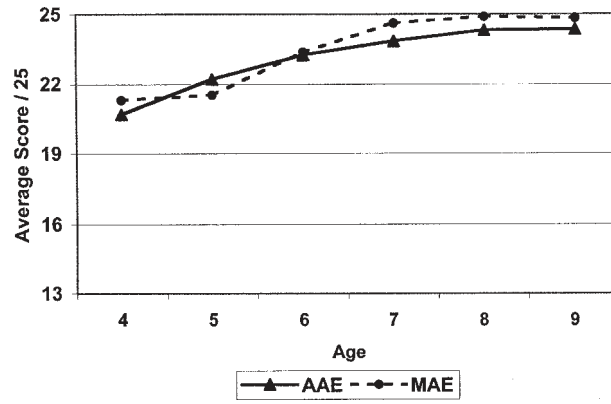


Figure 2 Phonology scores by Dialect (typically developing only. AAE, African American English; MAE, Mainstream American English.)

THE PHONOLOGY DOMAIN RESULTS

Developmental milestone testing of the Phonology probe showed its effectiveness in meeting the assessment objective; that is, the assessment of phonological disorders in a linguistically and culturally fair manner. With respect to dialect differences, there was no significant difference between AAE and MAE typically developing children across ages 4 to 9 (Fig. 2). As expected, young children made far more errors than older children. However, because of the nature of the task, differences among ages were not as great as those observed for the other language domains discussed in this issue, but the age differences were still statistically significant [Age, $F(5, 736) = 16.691$, $p < .0001$; Dialect, $F(1, 736) = .060$, $p = .806$; Age by Dialect, $F(5, 736) = .818$, $p = .537$].

An inspection of the distribution of mean scores across ages for impaired versus typically developing children shows an obvious separation across all ages except age 9 years (Fig. 3). This pattern is borne out in the statistics, which show strong effects of age and clinical status, and also an interaction between them: Age, $F(5, 1079) = 19.267$, $p < .0001$; Clinical Status, $F(1, 1079) = 180.40$, $p < .0001$; Age by Clinical Status, $F(5, 1079) = 3.807$, $p = .002$. This strong separation is also consistent within each dialect; that is, AAE-impaired children differ from AAE typically developing children (Fig. 4) (Age, $F(5, 670) = 8.088$, $p < .0001$; Clinical Status, $F(1, 670) = 117.08$, $p < .0001$; Age by Clinical Status, $F(5, 670) = 1.63$, $p = .146$], and the same results hold for MAE [Fig. 5, Age, $F(5, 380) = 7.655$, $p < .0001$; Clinical Status, $F(1, 380) = 46.231$, $p < .0001$;

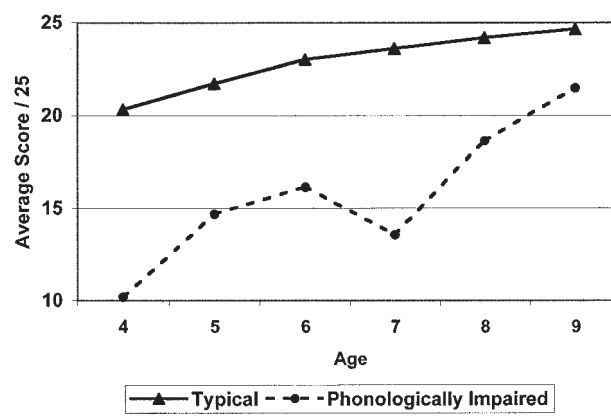


Figure 3 Phonology scores by Clinical Status (phonologically impaired versus not impaired).

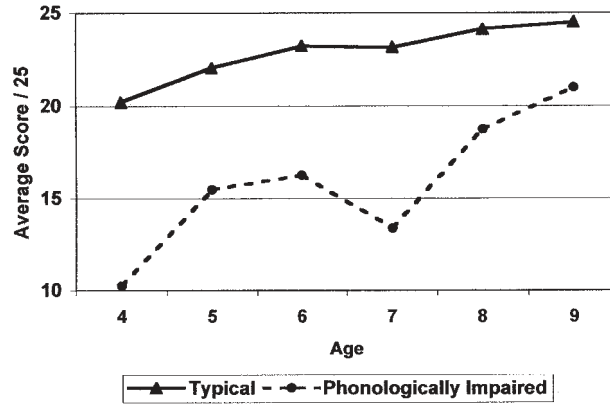


Figure 4 Phonology scores by Clinical Status (phonologically impaired versus not impaired), African American English (AAE) background only.

Age by Clinical Status, $F(5, 380) = 1.192$, $p = .313$).

IMPLICATIONS FOR FURTHER DIAGNOSIS

If a child's performance on the Phonology probe were to suggest a phonological problem, we believe that the child should be given an in-depth phonological evaluation. Because the phonology domain assesses clusters only, an examination of the full array of both singleton and cluster phonemes is recommended. This diagnosis along with error patterns on a Phonology probe such as the one we developed would form the basis for intervention strategies.

There are several traditional formats that are used to assess phonology and articulation in children.[32] These commonly include standardized tests that assess the full range of English consonants, connected speech samples, single-word corpus analysis, and stimulability testing, among others. The analysis applied to the child's phonological productions can be of several kinds, including segmental, natural process, and distinctive features. Design features reflected in the phonotactic and context characteristics of the Phonology domain can suggest a child's area(s) of weakness and motivate a particular format for testing and focus of analysis.

It is recommended that the clinician choose diagnostic procedures she or he prefers,

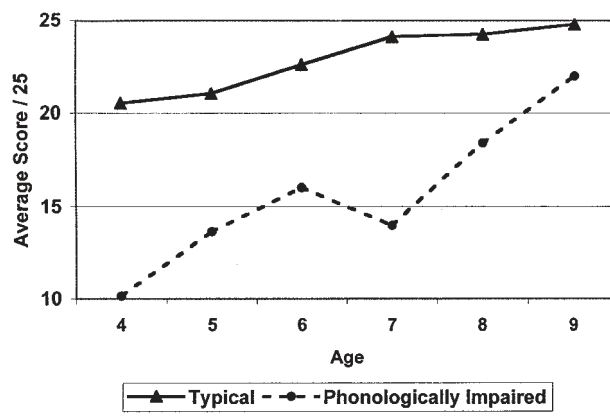


Figure 5 Phonology scores by Clinical Status (phonologically impaired versus not impaired), Mainstream American English (MAE) background only.

but that a battery employing several measures may be necessary to describe fully the nature and extent of a child's departure from age-appropriate phonological targets.

In addition, the planning, implementation, and interpretation of the diagnosis should be influenced by the child's dialect background. If the child is an MAE speaker one can employ traditional and "best practice" methods, which are familiar to speech-language pathologists (SLPs) and should represent the most current and common practices.[21,32–34]

Alternatively, if the child is not an MAE speaker, the examining clinician *must* be familiar with the phonological characteristics of the particular dialect spoken by the child. For example, if a child speaks AAE, there are certain predictable contexts in which consonants may be absent, such as in final positions of words and within certain consonant clusters. Specific to AAE, there are three types of phonological errors that must be taken into consideration.[35]

1. Type I refers to violations on consonants that are common or noncontrastive between AAE and MAE such as ([tat] for "cat").
2. Type II errors involve typical contrastive AAE patterns where an AAE speaker would produce [f] for a /θ/ in word final position [mouf], but an impaired child might produce [mous] (the [s] for /θ/ is a pattern of neither AAE nor MAE).
3. Type III errors involve impairment features that are qualitatively shared between the AAE and MAE but are quantitatively different. There is a higher probability in both dialects that in certain contexts a consonant cluster may be reduced, as when the cluster is followed by a consonant ("fast ball" → "fas ball") than when followed by vowel ("fast end"→ "fast end"). A type III error would involve a child producing a higher proportion of the cluster followed by the vowel than is expected for even an AAE speaker.

It is important to conduct an in-depth analysis to determine the extent to which the three types described above characterize the child's impairment. In characterizing the child's error types, as with the MAE child, one should employ traditional and "best practice" methods.

IMPLICATIONS FOR INTERVENTION STRATEGIES

Clinicians should employ best practice methods in devising and implementing phonology intervention strategies, which may include motoric approaches in treating phonemic segments and/or cognitive/linguistic approaches involving phonological processes or distinctive features.[21,32–34] Intervention strategies apply to both the selection of intervention target behaviors and the remediation procedures to address those target behaviors. Target behaviors should result from the in-depth diagnosis described above. Remediation methods typically involve both perception and production tasks.

The planning and implementation of intervention strategies should be influenced by the child's dialect background. If the child is not an MAE speaker, selection of target behaviors should first focus on Type I errors (those that do not contrast between AAE and MAE (such as [tat] for "cat"). Traditional remediation practices should prevail in addressing this kind of target behavior.

The next behaviors that should be targeted are Type III errors (those that generalize from a typical English cluster reduction ("fas ball") to an atypical one ("fast end"). Again, traditional remediation practices should prevail. The last target behaviors to be addressed would be Type II errors (those that violate the AAE rule, such as when a child fails to produce either the MAE ([θ]) or AAE pattern (/θ/ → [f] ("mouf"). This error pattern is more difficult to treat because a decision must be made as to whether the target should be the AAE pattern or the MAE pattern. It is possible that effective intervention for Type I and Type III could resolve Type II problems without direct remediation. However, the result is likely to be the preservation of the AAE pattern, in which case the clinician must be willing to accept the child's dialect pattern.

In the event that Type II errors are unaffected by intervention for Type I and III errors, it is recommended that a code-switching

model be implemented by stimulating the child to perceive and produce both the MAE and AAE patterns. It is possible that the child will continue to code switch, or choose to use the pattern most consistent with his or her community dialect, and either would be acceptable. Undoubtedly, this recommendation is controversial and SLPs may find it difficult to accept response patterns other than MAE. Nevertheless, code switching is a common and natural mode among AAE speakers and can be among the most realistic of clinical goals given the child's community dialect.

CONCLUSION

We believe that a dialect-sensitive phonological assessment should have the following characteristics:

1. It has a single scoring and test format regardless of a child's variety of English.
2. Both typically developing MAE and non-MAE speaking children perform equally well across ages.
3. AAE speakers are in no way penalized for speaking AAE.
4. A developmental age differential is evident among typically developing children of both MAE and non-MAE language backgrounds.
5. It is equally effective in differentiating phonologically impaired children from typically developing children of both MAE and non-MAE backgrounds.

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.

REFERENCE

1. Seymour HN, Seymour CM. A therapeutic model for communicative disorders among Black English speaking children. J Speech Hear Disord 1977;42: 247–256
2. Seymour HN. The challenge of language assessment for African American English-speaking children: a historical perspective. Semin Speech Lang 2004;25:3–12
3. Wolfram W, Fasold RW. The Study of Social Dialects in American English. Englewood Cliffs, NJ: Prentice Hall; 1974
4. Green L. African American English: A Linguistic Introduction. Cambridge, MA: Cambridge University Press; 2002
5. Stoel-Gammon C, Dunn C. Normal and Disordered Phonology in Children. Austin, TX: Pro-Ed; 1985
6. Seymour H, Seymour C. Black English and Standard American English contrasts in consonantal development of four- and five-year-old children. J Speech Hear Disord 1981;46:276–280
7. Haynes W, Moran M. A cross-sectional developmental study of final consonant production in southern black children from preschool through third grade. Lang Speech Hear Serv Sch 1989;20: 400–406
8. Stockman I. Phonological development and disorders in African American children. In: Kamhi

AG, Pollock KE, Harris JL, eds. Communication Development and Disorders in African American Children: Research, Assessment, and Intervention. Baltimore: Paul Brookes; 1996:117–153
9. Stockman I. Variable word initial and medial consonant relationships in children's speech sound articulation. Percept Mot Skills 1993;76:675–689
10. Abdulkarim L, Bryant T, Seymour HN, Pearson BZ. Final consonants in child African American English phonology. Poster presented at: the Annual Meeting of the American Speech-Language-Hearing Association, San Francisco, November 18–21, 1999
11. Abdulkarim L, Bryant T, Seymour HN, Pearson BZ. Consonant clusters in child African American English phonology. Poster presented at: the Annual Meeting of the American Speech-Language-Hearing Association, San Francisco, 1999
12. Pearson BZ. Theoretical and empirical bases for dialect-neutral language assessment: contributions from theoretical and applied linguistics to communication disorders. Semin Speech Lang 2004; 25:13–26
13. Ciolli L, Seymour HN. Dialect identification versus evaluation of risk in language screening. Semin Speech Lang 2004;25:33–40
14. Templin MC. Certain Language Skills in Children. Minneapolis, MN: University of Minnesota Press; 1957
15. Sander EK. When are speech sounds learned? J Speech Hear Disord 1972;37:55–63
16. Prather E, Hedrick D, Kern C. Articulation development in children aged two to four years. J Speech Hear Disord 1975;40:179–191
17. Yavas M. Phonology Development and Disorders. San Diego: Singular; 1999
18. Oller DK. Simplification as the goal of phonological processes in child speech. Lang Learn 1975; 24:299–303
19. Macken MA, Ferguson CA. Phonological universals of language acquisition. In: Winitz HB, eds. Native Language and Foreign Language Acquisition. New York: New York Academy of Sciences; 1981
20. Vihman MM. Later phonological development. In: Bernthal J, Bankson N, eds. Articulation and Phonological Disorders, 4th ed. Boston: Allyn & Bacon; 1998:113–147
21. Velleman SL. Making Phonology Functional: What Do I Do First? Boston: Butterworth-Heinemann; 1998
22. Ingram D. Phonological Disability in Children. New York: American Elsevier; 1976
23. Gerken LA. A metrical template account of children's weak syllable omissions from multisyllabic words. J Child Lang 1994;21:565–584
24. Kehoe M, Stoel-Gammon C. The acquisition of prosodic structure: an investigation of current accounts of children's prosodic development. Language 1997;73:113–144
25. Velleman SL, Shriberg LD. Metrical analysis of the speech of children with suspected developmental apraxia of speech and inappropriate stress. J Speech Lang Hear Res 1999;42:1444–1460
26. Olmsted D. Out of the Mouth of Babes; Earliest Stages in Language Learning. The Hague: Mouton; 1972
27. Dubois E, Bernthal J. A comparison of three methods for obtaining articulatory responses. J Speech Hear Disord 1978;43:295–305
28. Seymour H, Green L, Huntley R. Phonological patterns in the conversational speech of African-American Children. Poster presented at: the Annual Meeting of the American Speech-Language-Hearing Association, Atlanta; November 22–25, 1991
29. Morrison JA, Shriberg LD. Articulation testing versus conversational speech sampling. J Speech Hear Res 1992;35:259–273
30. Moran M. Final consonant deletion in African American children speaking Black English: a closer look. Lang Speech Hear Serv Sch 1993;24:161–166
31. Bryant T, Abdulkarim L, Velleman SL, Seymour H. Predictors of coda consonant cluster reduction in African American English. Paper presented at: the Meeting of the American Speech Language and Hearing Association, Atlanta, GA, November 21–24, 2002
32. Bernthal JE, Bankson NW. Articulation and Phonological Disorders, 4th ed. Boston: Allyn & Bacon; 1998
33. Bleile KM. Manual of Articulation and Phonological Disorders. San Diego: Singular; 1995
34. Shriberg LD, Kent RD. Clinical Phonetics, 2nd ed. Boston: Allyn & Bacon; 1995
35. Wolfram W. The phonology of a sociocultural variety: the case of African American Vernacular English. In: Bernthal JE, Bankson NW, eds. Child Phonology: Characteristics, Assessment, and Intervention with Special Populations. New York: Thieme; 1994:227–244

Distinguishing Dialect and Development from Disorder: Case Studies

Barbara Zurer Pearson, Ph.D.,[1] and Lois Ciolli, M.A., C.C.C.-S.L.P.[2]

ABSTRACT

Seven case studies are presented to illustrate how the tasks in the proposed test battery provide the tools to distinguish language differences due to development or dialect from true signs of delay or disorder. The case studies exemplify different combinations of language strengths and weakness found among participants of extensive field research in the age range from 4 to 9 years. Special attention is paid to certain aspects of language development, such as time clauses or double *wh*-questions, in which dialect features play no role. All levels of performance, from the highest to the lowest, are found in children who are speakers of Mainstream American English (MAE) and the same is true for those whose language patterns indicate a strong variation from MAE.

KEYWORDS: Profile, path of progress, subdomain, item types, strength, weakness, interpretation

Learning Outcomes: As a result of this activity, the participant will be able to identify (1) how breaking down the items on the proposed diagnostic test into subtypes helps to profile the child's strengths and weaknesses, and (2) why it is vital to consider all three factors—development, dialect, and disorder—when evaluating a child's language.

First and foremost, the goal for a dialect-sensitive language assessment is to help clinicians identify the signs of language disorder in children without penalizing them for dialect features or typical linguistic development. Some children will exhibit signs of all three factors (i.e., disorder, dialect, and development); others will exhibit just one or two signs in different combinations. The ideal assessment helps clinicians recognize when dialect features are present alone or in conjunction with signs of delayed or disordered development, and also when delay or disorder is apparent, but dialect features are not. As with all assessment

Evaluating Language Variation: Distinguishing Dialect and Development from Disorder; Editors in Chief, Nancy Helm-Estabrooks, Sc.D., 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., Research Associate, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, MA 01003. E-mail: bpearson@comdis.umass.edu. [1]Research Associate, Department of Communication Disorders, University of Massachusetts Amherst, Amherst, Massachusetts; [2]The Psychological Corporation, Harcourt Assessment, Inc., San Antonio Texas. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01,101,112,ftx,en;ssl00186x.

instruments, the results need to be considered along with all other information gathered about the child.

In this article, we explore the profiles of seven children from our field research to illustrate how the tasks in the proposed test battery provide the tools to make the required distinctions. The case study children were chosen to exemplify different combinations of strengths and weaknesses and different levels of dialect usage at different points in the age range 4 to 9 years. Two 4-year-old children are used to contrast typical versus delayed development. In addition, both speak with "Some Variation from Mainstream American English (MAE)," but contrast with respect to how the results of their Language Variation Status should be interpreted. The expectations for performance of a 4-year-old child are contrasted with those for a 9-year-old child, who shows relatively fewer errors, but more serious risk and/or disorder. Two 6-year-old children also illustrate the contrast between a diagnosis of disorder versus no disorder, and in addition show how our proposed assessment measure may reveal linguistic strengths in children in whom such strengths are less apparent—in the quiet child and in the child with a severe phonological disorder. Finally, two 7-year-old children illustrate the contrast between a non-MAE speaker with no disorder and a disordered child who speaks MAE. Ironically, the former is receiving language services, whereas the latter is not.

In these case studies, the dialect or Language Variation Status designations for each child are derived directly from our screening items, and further illustrated with examples taken from the child's answers to other items of the comprehensive language test in which the children's short-sentence answers were written verbatim on the record form by the examiner. The diagnostic risk designations come from the child's score on the second set of screening items.

GENERAL GUIDE TO GROWTH BY SUBDOMAIN

Summarizing across the tasks in the diagnostic portion of the screening test and the four domains of the proposed comprehensive language test, there are 14 item-types represented by the names of the subdomains on the record form: for example, fast mapping, short narrative, *wh*-question comprehension, and so on (see Fig. 1 in Seymour and Pearson[1]) Those subdomains are further subdivided at the item-type level into target skills or concepts. Thus, the Syntax domain *wh*-questions are of three subtypes: double *wh*-questions, embedded false clauses, and barrier questions; the Articles subdomain has items of two subtypes, definite ("the") and indefinite ("a"). The Semantics domain fast mapping items can be viewed as two subtypes, real verbs and novel verbs, or as three subtypes across real and novel, according to whether the item involves transitive, transfer, or complement structures. In all, there are over 30 item-subtypes. (The subtypes and their rationales are discussed individually in the articles on the individual domains (Roeper,[2] P. de Villiers,[3] J. de Villiers,[4] and Seymour[5].)

For the various subdomains and item subtypes, there is a general path of progress, such as the ones outlined in Table 1 for *wh*-questions, Table 2 for short narratives, and Table 3 for nonword repetition (from the screener, part 2). There are also developmental graphs in the domain articles that provide similar growth ranges for many of the constructs in the various subdomains.

These growth ranges represent an approximate timeline to help evaluate the children's responses from a qualitative point of view. However, one cannot say a child is "functioning like a 4 year old" because he or she gave two of two medial answers, corresponding to the description for that construct under 4-year-old children in Table 1. One must take several factors and more than one subdomain into account to make such a judgment. However, the tables indicate the nature of our expectation of progress as the child grows. Even though most 4-year-old children make errors in all four areas noted, only the weakest ones fail at *all* of them. For example, if a 4-year-old child passes the false clause items and the relative clause items, but makes medial and singleton errors on the other questions (Roeper[2]), that would be an average or above-average four-year-old per-

Table 1 Approximate Growth Stages for *Wh*-Comprehension Item-Types

Wh-Comprehension	TD; LI	4 Years	5 Years	6 Years	7 Years	8 Years	9 Years
Wh-paired lists	TD	40% have singleton errors	30% singleton errors	20% singleton errors		Rare singletons, less than 10%	
	LI	Almost no paired-exhaustive answers				LI fewer than 70% paired exhaustive answers; still almost 30% singletons	
Wh-barriers	TD	Miss 2 of 2 OK, (almost 1/2 have medial errors)	Miss 1 OK	Only 20% medial errors	Miss 1 rare	Medial errors 5 to 10%	
	LI					LI fewer than 50% correct (SD) answers Equal medial and ''other off-target'' errors	
Relative clause barriers	TD	Only 60% give matrix answer	70% TD give matrix answer			Almost no errors	
		Barrier error and ''other irrelevant'' about equal		Fewer than 10% barrier errors			
	LI				LI still 20% barrier errors		
False clause items	TD	60% give 2-clause answer		80% give 2-clause answer		90% + give 2-clause answer	
	LI				LI still give 30% 1-clause answers: many other off-target answers		

LI, language impaired; TD, typically developing.

Table 2 Growth Stages for short Narratives

Temporal Links	TD; LI	4 Years	5 Years	6 Years	7 Years	8 Years	9 Years
No time words		About half of the children use no time words		About 1 in 10 children use no time words	Almost no children with children no time words		
Sequencers (*then*, and *then*)		At all ages about half of the children use sequencers (even LI children from age 5 on)					
Temporal clauses with conjunctions and verbs	TD	Very few have them	About 1 in 5 children has at least one clause		About half of the children use them		
	LI		About one fourth use time clauses				
Theory of Mind References (at Least One)		**4**	**5**	**6**	**7**	**8**	**9**
References to actions only		About three fourths of the children	About 1 of 2 children	About 1 in 4 children			Very rare
References to desires or intentions			About 1 in 3 children				About 1 in 2 children
Mental verbs and explanation of actions	TD	Very rare	About 1 in 10 children has at least one	About 1 in 5 children		About 2 in 5 children	About 1 in 2 children
	LI		About 1 in 20 children uses at least one mental verb	About 1 in 10 children			LI catch up to TD

LI, language impaired; TD, typically developing.

Table 3 Expectation for Errors in Nonword Repetition

4-5 Years	6-7 Years	8-9 Years
Expect 1 to 2 errors on 3- and 4-syllable prompts	Expect 1 error on 4-syllable prompt	Expect few errors

formance for the *wh*-question subdomain. A 6-year-old child with the same set of answers would be below average, and if, for example, she or he gave all four double *wh*-questions as singleton answers, that alone would point to a low average or weak performance in that subdomain because 6-year-old children are expected to get at least some of the double *wh*-questions correct. For a 9-year-old child, a single nonexhaustive (singleton) answer would be a warning signal. Keep in mind, however, that Table 1 presents only the average based on the sample reported by Roeper[2]; it does not present ranges of responses by age.

THE CASE STUDIES

The profile for a case study child makes reference first to an overall evaluation and a summary of scores by domain. To characterize what makes the child average or below or above in a domain, we comment on only some subdomains and item types—generally those that stand out as different from the kind of answers given by typically developing children of the same age or as different from other parts of the child's own performance. The case studies illustrate how scores on all four domains interact and help the clinician make recommendations for services (see discussion in Seymour and Pearson[1]).

DISTINGUISHING DEVELOPMENT FROM DIALECT

Case Study 1: Charnelle (Age 4 years, 2 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: Yes
Dialectal: Yes
Disordered: No

Charnelle (Table 4) is a typically developing 4-year-old African American child whose speech and language was screened upon her entrance into prekindergarten in the neighborhood public school. She is the youngest member of a large family in the midwest. Her parents are high school graduates and both are employed outside the home.

DEVELOPMENTAL VARIATION

As presented in Table 4, Charnelle scored in the "Some Variation from MAE" range on the proposed Language Variation Status items and showed medium risk on the diagnostic portion of the screening test, so the dialect-sensitive comprehensive language test was the recommended follow-up.

As expected because of her young age, Charnelle shows a mixture of strengths and weaknesses. These developmental variations

Table 4 Charnelle (4 years, 2 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Some variation from MAE	Medium risk for disorder	17 Average	12 Average	24 Average	20 Average

MAE, Mainstream American English.

are mixed with dialectal variations due to the influence of the language spoken most often in her home and community (i.e., African American English [AAE]). For example, she produced the "th" sound as [f], [v], or [d] in predictable AAE contexts (e.g., "mouf" and "wiv" for "with," and "anoder" for "another"). Like many young MAE and AAE 4-year-old children, Charnelle finds "th" difficult to produce in other contexts not associated with dialect differences, such as "frowing" for "throwing." She also still is having trouble pronouncing the /r/ and /l/ sounds in words such as "play" ("pway") and "cracker" ("kwacker"), so her speech contains some typical developmental errors.

This combination of development and dialect also is seen in her morphosyntax items: she gave several MAE responses in the identifier section (part 1) of the screening items, in possessive and auxiliary verb items, and some predictable dialect responses as well. This mixture of MAE/non-MAE responses earned her a dialect score of "Some (but not strong) Variation from MAE." When she was responding to items on the comprehensive language test, the examiner noted that Charnelle used single negation (e.g., "has no shoes"; an MAE feature) three times and multiple negation (e.g., "don't have no shoes"; a feature characteristic of AAE) only once. Throughout both the screening and diagnostic tests, Charnelle gave many responses that could only be classified as "other;" most of them were incomplete answers or answers that indicated that she did not understand the target of the items. In pragmatics question asking, like many of the 4-year-old children, she had difficulty suppressing the temptation to *answer* rather than *ask* questions, and she reached her ceiling early.

Despite such answers that show her developing status, in several subdomains of the comprehensive language test, she responded with surprising maturity. In the short narratives, Charnelle contrasted the characters and used time sequencers; she understood the importance of "thought balloons" in the pictures and reported what the characters were dreaming about and what they wanted, instead of focusing only on actions or descriptions of the pictures, which is more typical of a child her age. In *wh*-questions, she also gave answers unusual at her age: she got all three double *wh*-questions correct, giving exhaustive, paired responses ("she caught the crab with a net and a fish with her pole," [see Roeper[2]]), and she responded correctly to the complex question about what "the sister *said* she brought." In the passive and quantifier subdomains, she also had a combination of both very strong and very weak performances.

SUMMARY

Both dialect and developmental influences are seen in Charnelle's test performance. Although she appeared to be at some risk for a language disorder (i.e., she showed medium risk on the screener), she passed the comprehensive test, scoring in the average range in all four domains. She illustrates how the two tests taken together help distinguish which features are due to development and dialect. In Charnelle's case, a diagnosis of disorder does not appear warranted.

Case Study 2: Carla (Age 4 years, 0 months)

PROBABLE CONTRIBUTING FACTORS:

Developmental: Yes
Dialectal: No
Disordered: Yes

Carla (Table 5) is a non-Hispanic white child in a Head Start program in the northeastern United States. She was screened as part of Head Start's routine program. Both of her parents have completed high school. Although the mother works part time outside the home, she also volunteers 1 day a week at Carla's preschool.

Like Charnelle, Carla's Language Variation Status is in the "Some Variation from MAE" range. Because there is no indication that she is from a minority community, there is no reason to attribute this score to dialect background. It is more likely the consequence of delayed development or disorder. That is, her variation from MAE may be caused by more off-target responses as well as some remnants of immature language in her speech. Her

Table 5 Carla (4 years, 0 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Some variation from MAE	Highest	15 Low	5 Weakness	18 Low	18 Average

MAE, Mainstream American English.

phonology, although in the average range, still has several poorly articulated consonants, mostly distortions. We also see in her verbatim responses terms such as "has an owee" and "a boo-boo." However it is her high Diagnostic Risk Status that suggests that her overall language development is slower than that of her peers. Carla's performance on the comprehensive language test shows her to be in the low average and weakness score category in three of the four language domains. In a few subdomains she is showing age-appropriate progress, but there are several in which she shows serious problems. In no subdomain does she show strength.

The examiner noted that although Carla was pleasant and cooperative in the test session, she had difficulty with almost all of the directions. In question asking, when told to "Ask me a 'what' question," she merely repeated the words, "a what question;" similarly, she repeated "a who question." In short narratives, her answers to the follow-up questions were off-target. For example, when asked, "Why was she looking *there*?" (with emphasis on "there"), Carla did not focus on the character's choice of a place to look as the contrastive stress directs the child to do. Instead she told the examiner why the girl wanted the item she was looking for ("'cause she's going to throw it in the garbage"). Although her response is not completely unrelated, there is nothing in the pictures to suggest such an idea, so it looks like she is not interpreting the prompt and is falling back on her nonverbal interpretations of the pictures.

Case Study 3: Dominique (Age 9 years, 11 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: No
Dialectal: Yes
Disordered: Yes

Dominique (Table 6) is an African American 9-year-old child in third grade in a small town in the south. He has been receiving language services for 1 year on the basis of previous testing. Because there was some question as to the validity of the previous testing, Dominique was given the dialect-sensitive test battery described in this issue (Screening and Comprehensive Language Test) to confirm his continued eligibility.

Because he is 9, at the upper end of the age range, he gets many responses correct. A few "red flags," or crucial errors, however, are enough to confirm his high risk status on

Table 6 Dominique (9 years, 11 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Strong variation from MAE	Medium to high risk for disorder	25 Weakness	21 Average	33 Weakness	25 Strength

MAE, Mainstream American English.

the screening test and predict his areas of "weakness" on the comprehensive language test. For example, Dominique's medial answer to a *wh*-barrier item raises a question. Although some typically developing 9-year-old children (~10%) are still consolidating their ability to interpret those questions and will miss one from time to time, Dominique missed *all* of the double *wh*-questions. Fewer than 1 in 10 typically developing children of his age would give even one singleton answer (see Table 1). Another strong indicator is found in the quantifier subdomain on one of the *every* items. Dominique's answer shows he extends *every* to all of the elements in the picture, rejecting the statement about "every cat" because there's a rabbit doing something else. This is a very common (40%) error among 5-year-old children, but this type of error occurs much less frequently by age 8 years—except among language-impaired children (Table 6).

The examiner also noted other behavioral signs of Dominique's language difficulties. Instead of answering, he repeated one of the prompts in the preposition contrasts items, and three times he repeated his first answer to verb contrast items, despite being prompted to "tell me a different word, not one you've already said." This indicates that he is having difficulty organizing his lexicon so that it can be used flexibly and appropriately.

SIX-YEAR-OLD CHILDREN: DISCOVERING COMPETENCIES WITH THE PROPOSED TESTS

To be average at age 6 requires a higher percentage of correct answers than at age 4, but there is still room for some unevenness in a child's performance. The picture in all domains taken together must be considered. Failure in one domain gives the clinician or teacher suggestions for areas that may need special attention, but weakness in one domain does not indicate, by itself, disorder.

Case Study 4: Serina (Age 6 years, 0 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: No
Dialectal: Yes
Disordered: No

Serina (Table 7) is an African American 6-year-old child in a suburban public school in a racially mixed neighborhood in the western region of the United States. Her father has a college degree and her mother has some college background as well. According to the first set of screening items, Serina speaks with a strong variation from MAE. Although her responses on part II put her in a low-risk category, she is one of the youngest children in her first grade class and she rarely speaks up in class. Because of her teacher's concern, she was evaluated further with the comprehensive language test.

Serina failed (scored in the weakness category) in the Semantics domain. Indeed, her weakness in three of the four subdomains of Semantics may have been what alerted her teacher to a possible problem. However, in all other areas tested, Serina's scores were average or above and provide evidence of her adequate language abilities overall.

Serina's lowest score relative to her peers was in verb contrasts, where she missed 7 of 10 items. However, she did not make the most immature kinds of errors, such as repeating a prompt or giving the same answer to two

Table 7 Serina (6 years, 0 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Strong variation from MAE	Lowest	26 Average	21 Strength	24 Weakness	22 Average

MAE, Mainstream American English.

different prompts. Examination of the answers she gave showed that she had difficulty matching the contrast of the prompt, giving too specific a word three times and too general a word three other times. In quantifiers she also had difficulty with comprehending a precise word meaning. She showed that she did not yet understand the meaning of *every*, much less have an appreciation of its special properties (as discussed by J. de Villiers[4]). In her fast mapping, she did better with the real verbs than the novel verbs, indicating that she may be relying on lexical cues, and not taking as much information as she could from the structure of the sentences she hears.

In contrast to her performance in Semantics, Serina scored above average in the Pragmatics domain. She was particularly good in recognizing the false belief of the character in the short narrative and in using the language of thought to describe the motivations for events in the story. In her question asking, she needed a second prompt for almost all of the items, but with a little extra guidance demonstrated her competence at recognizing what she did not know and asking the appropriate question to learn it. In both of these subdomains she demonstrated not only that she understood what she was shown, but also that she was adept at using what was stated to find out more.

Case Study 5: Leon (Age 6 years, 4 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: Yes
Dialectal: Yes
Disordered: Yes (but with unexpected strengths)

This case study illustrates a situation in which the child's unintelligibility makes accurate diagnosis difficult. Leon (Table 8) is a 6-year-old African American child who lives with his mother in a northern city. Leon's mother has less than a high school education. Like Serina, Leon also failed the Semantics domain and excelled in the Pragmatics domain, but there the similarity ends. He is repeating kindergarten, and both he and his teacher are having a difficult time. Leon's speech errors are so severe that he is almost impossible to understand. The screener confirmed that Leon substitutes final consonants and simplifies clusters in a way consistent with an AAE dialect, but the results of the Phonology domain show that his variation in his speech goes far beyond dialect difference and that he would profit from a full phonological evaluation. Almost one third of the words he produced on the Phonology domain had at least one consonant that was so distorted that the examiner could not transcribe it. Furthermore, he simplified initial clusters, as a younger child might, but in addition, he omitted *initial* consonants. This pattern is very damaging to intelligibility and is not characteristic of either dialect differences or typical developmental patterns.

Given the obvious nature of his problem, Leon had already been diagnosed with a phonological and language disorder before he was given our assessment items, and he was already scheduled to begin receiving them. The speech-language pathologist gave him the experimental probes to help identify therapy goals for him. In Leon's case, the innovative items were useful in showing unexpected strengths.

On part II of the screening items, Leon omitted almost all of the past tense copulas ("was") and did not even attempt the three- and four-syllable prompts in the nonword

Table 8 Leon (6 years, 4 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Strong variation from MAE	Highest	20 Weakness	20 Strength	20 Weakness	5 Weakness

MAE, Mainstream American English.

repetitions. On the comprehensive language test he showed very weak performance on articles, preposition contrasts, and fast mapping. In all of these areas, successful performance depends on good comprehension of short, unstressed elements in the stream of speech. To succeed at fast mapping, for example, the child must attend closely to the endings of words (such as the *-er*, *-ing*, and *-ible* endings). These all appear to be compromised in Leon's language system.

However, as noted above, Leon showed strength in Pragmatics. In particular, he got all but the double *wh*-question correct in question asking, which is unusually good for a 6-year-old child. Although his utterances lacked the unstressed elements, which caused him difficulty in other parts of the test (such as, "Why she so mad?" and "Who cake there?" for "Who*se* cake *is in* there?"), he was above average in his understanding of the communicative needs in the scenarios presented. Similarly, in the short narrative, he recognized and articulated that the cake was moved and so the character who did not see it being moved would not know where it was. In the field research, only about one in five 6-year-old children showed this high level of understanding. Finally, his *wh*-question comprehension was also above the expected level for typically developing children his age. He gave a singleton answer one time where a paired exhaustive answer was expected (see Roeper[2]), but that was his only error. He got the other paired exhaustive answer right and both barrier questions. Consistent with his mature "theory of mind" answers in short narrative, Leon also correctly responded to the item that assesses a child's comprehension of complements with a false clause (he answered what she *said* she brought, not what she brought).

Clearly, Leon will profit from intensive speech and language therapy, but it may have a positive effect for his teachers to realize that despite the fact that he cannot make himself understood, he is quite advanced in his understanding of what others say to him.

INDEPENDENCE OF LANGUAGE VARIATION AND DIAGNOSTIC STATUS

It is especially difficult to be aware of linguistic strengths when the child uses stigmatized (i.e., non-MAE) word forms, such as double-marked past tense ("he holded it") or an invariant case of a pronoun (as in "them boys" or "her going home"). These occur variably and are not predictable dialect forms, so their presence in a child's speech may be problematic. In cases such as these, the dialect-sensitive language test can highlight other sophisticated knowledge of the child that may not be evident in day-to-day conversation.

Case Study 6: Dejean (Age 7 years, 1 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: No
Dialectal: Yes
Disordered: No

Dejean (Table 9) is a 7-year-old child from the south currently receiving language services in his school. His Screener Language Variation Status registers as "Some Variation from MAE," which alerts us to the appearance in his speech of predictable dialect forms that occur variably. That is, on the screener he used some features of MAE (e.g., "he has," "he talk*s*"), and some features of AAE (e.g., "he don't," "he climb"). On the phonology items on the screener he produced some MAE forms

Table 9 Dejean (7 years, 1 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Some variation	Lowest	32 Strength	22 Strength	39 Strength	25 Strength

mixed with non-MAE forms. Indeed, in the responses that the examiner captured verbatim, in addition to many predictable dialect forms, we see elements that are immature, but not characteristic of a particular dialect (e.g., "them can't," "he hided it"); so he appears to be in the process of maturing as well as possibly learning to code-switch, but his learning is still incomplete.

In contrast to the uncertainty about his dialect status, he shows the lowest risk level on the diagnostic portion of the screener and his domain scores on the comprehensive language test are uniformly high. In fact, his domain scores are high enough to be considered "strengths" even if he were 2 years older. On the short narrative follow-up questions, for example, he made references to the characters' thoughts, putting his answers in the most mature category. (e.g., "She's thinking about the cake," "He didn't know his sister put it in the refrigerator"). Most children of his age would pass a theory of mind test (see P. de Villiers[3]), but fewer than half realize how important it is to include characters' thoughts and motivations in their stories and report them spontaneously, as Dejean did (Table 2).

In his *wh*-question asking, Dejean most often gave the right answer at the first prompt. In the three cases when he did not, he responded effectively to the examiner's prompt, even when it was indirect, as in recognizing, for example, that "for a reason" requires a *why*-question. Still more impressive is his ability on the fast mapping task. He did as well on the novel verbs as with the real verbs, demonstrating that he was truly processing the inflections *-er, -ing*, and other syntactic cues and did not rely on stored knowledge.

Case Study 7: Nia (Age 7 years, 8 months)

PROBABLE CONTRIBUTING FACTORS

Developmental: Yes
Dialectal: No
Disordered: Yes

In contrast to Dejean, Nia (Table 10)—a 7-year-old African American child also from the south—uses enough MAE features to be categorized on the screener as an MAE speaker. She performed perfectly on the Phonology domain of the comprehensive language test and did very well at the nonword repetition task, missing only 1 of 6, so we have an impression of a clear speaker with a good phonological memory. Her use of MAE surface inflections, however, does not mean that she commands the deep syntactic principles of the language.

In fact, Nia's diagnostic score for the screener indicates the highest level of risk. She used auxiliary *is* or *are* where a non-MAE speaker would generally omit them, but she used them in ways that revealed serious grammatical problems. For example, in the diagnostic portion of the screener, she used the present tense *are* (e.g., "the dogs are sleeping,") when the past tense verb *were* was called for. When asking questions in the pragmatics question asking items, she used the MAE *is*-copula with inversion (e.g., "What is it?"), but the poorly focused question failed to make reference to the action and objects highlighted in the prompt and the picture. Another of her responses was a well-formed, but irrelevant question (i.e., "Who made the cookies?")

Nia scored below the expectation for her age in three domains, most especially in Syntax.

Table 10 Nia (7 years, 8 months)—Summary of Scores

Screening Items		Comprehensive Language Test			
Language Variation Status	Diagnostic Risk Status	Syntax	Pragmatics	Semantics	Phonology
Some variation	Highest	18 Weakness	8 Low average	27 Low average	25 Strength

On *wh*-comprehension items she produced four of the particular "red flag" responses that show difficulty with complex sentences, and four nonexhaustive responses to the double *wh*-items. For example, instead of answering both *wh*-items in "Who ate what?", she responded only to the "what" ("the pizza and the spaghetti") and neglected the "who" in all such items. In the passive items, she responded like a younger child, both in the number and kinds of errors, understanding the prompt as an active sentence three times and as a completed action rather than an ongoing event three times. Similarly, in the fast mapping, she did much better with the real verbs than the novel verbs, showing that she was not using the grammatical information present in the sentence well enough to answer correctly when there were no lexical clues.

This illustration shows how grammatical knowledge and dialect are independent of each other. Even among African American children, all levels of performance from the highest to the lowest are found in children who speak the mainstream dialect, and the same is true for those who show a strong variation from MAE.

SUMMARY

As can be seen by the case studies, the variations in children's language can be explained by the presence or absence of various factors: typical developmental patterns that reflect a child's maturation toward an adult grammar; dialectal patterns that reflect a child's cultural and linguistic community; and in some cases, patterns that indicate slow, delayed, or disordered development. When several of these factors are involved, as they often are, the diagnostic process is more complicated. Therefore, it is most important to have an assessment battery, such as the one proposed in this issue, that can help sort out the factors involved in the variations to ensure the appropriate placement of children in special programs.

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 reseach 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. Seymour HN, Pearson ZP. Steps in designing and implementing an innovative assessment instrument. Semin Speech Lang 2004;25:27–32
2. Roeper T. Diagnosing language variations: underlying principles for syntactic assessment. Semin Speech Lang 2004;25:41–56
3. de Villers P. Assessing pragmatic skills in elicited production. Semin Speech Lang 2004;25:57–72
4. de Villers JG. Cultural and linguistic fairness in the assessment of semantics. Semin Speech Lang 2004; 25:73–90
5. Seymour HN. A noncontrastive model for assessment of phonology. Semin Speech Lang 2004;25: 91–100

Conclusions, Future Directions, and Implications for Remediation

Harry N. Seymour, Ph.D.,[1] Thomas Roeper, Ph.D.,[2] and Jill G. de Villiers, Ph.D.[3]

Learning Outcomes: As a result of this activity, the participant will be able to (1) identify how the proposed assessment achieved equity between African American English (AAE) and Mainstream American English (MAE) speakers, and (2) identify areas of future research.

The articles in this issue have provided a demonstration that achieving equity in language testing for speakers of different varieties of American English is within our reach. By avoiding areas of the language where dialect differences loom large, it has proven possible to tap children's knowledge of deeper aspects of language where disorder makes itself most apparent. In phonology, we avoid just those distinctive properties that identify a dialect speaker and assess other properties. In morphosyntax, we choose the inflections that remain obligatory for speakers of variants of American English. In syntax, we examine the deep properties of grammar that are fundamental to building a full understanding of language. In semantics, we choose to focus on basic processes and organization, avoiding the problem of different cultural settings. In pragmatics, we choose those aspects that can be evaluated in such a test—those that are neutral across the cultures and essential for schooling and literacy. In every case we find tests that can differentiate solidly and across this wide age range between typically developing children of either language background and those who exhibit language-learning difficulties.

When the project began, we did not know if this would really work. For example, it may have turned out that children described as specific language impaired (SLI) have problems *only* with the superficial morphology, and then the problem would have been a much less tractable one. It might have proved necessary to design subtle tests tapping the particular features of the various dialects of English, such as negative concord or aspectual morphology in African American English, to reveal problems for speakers who have SLI. It is our intention to continue research on that question, but the test would then have become targeted only to speakers of that variant of English. We believe that the reception for a test such as the one described in this issue is strengthened by its dialect-neutral application. Nonetheless, it is vital to have more linguistic work on how the variants of English develop, and the stages that children go through in arriving at the special properties they have. As discussed in the article by Pearson,[1] a small difference in an item such as number agreement or aspect might have widespread ramifications throughout the system that we cannot yet fully imagine. This could have important consequences for how

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: Jill G. de Villiers, Ph.D., Professor of Psychology and Philosophy, Smith College, Northampton, NA 01063. E-mail: jdevil@smith.edu. [1]Professor Emeritus, Department of Communication Disorders and [2]Department of Linguistics, University of Massachusetts Amherst, Amherst, Massachusetts; Smith College, Northampton, Massachusetts. Copyright © 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel: +1(212) 584-4662. 0734-0478,p;2004,25,01,113,116,ftx,en;ssl00187x.

the child's language is assessed, and particularly on the direction of clinical intervention.

Despite the success of the field testing, there are several areas that remain essential for the test to contribute to a new era of collaborative work between linguists and specialists in communication disorders. When we designed the tests, we had in mind to use the different items as a way to diagnose truly where a child might have missed a step in building the grammar, or mis-set a parameter, or might be missing a crucial ingredient that permeated several domains. This remains on our current research agenda, and we are making discoveries as we mine the rich dataset. The goal is to examine each item and categorize success or failure on it as a function of its properties—both performance demands, such as *memory* or *attention*, and linguistic, such as *recursion*, *agreement*, or *specification*. If the pattern of failures of a particular child can then be entered into a relational database, it is possible that the nature of the disorder can be revealed by determining what the items have in common. In the future, a clinician may be able to enter a child's pattern of mistakes and the computer would suggest a profile of difficulties. We have alluded to some such examples in our particular articles, but this is an ongoing endeavor, and it will involve deep linguistic insight as well as methodological sensitivity. Previous assessment tests have nothing like this level of detail. The generalizations about a child's language problem found in most tests of language fail to offer solutions to the speech-language pathologist in terms of the practice of remediation. The aim of the present test is higher, but the problems we identify must then be addressed by discussion of remediation.

Contemporary work on language acquisition from a linguistic perspective has been the inspiration of large parts of the current test, but it offers relatively little insight into the process by which children arrive at their rich knowledge. Yet the process is precisely what needs to be understood to offer support for remediation of the children's difficulties. There are two different meanings of *process* that are easily confused. Good psycholinguistic work has revealed the intricate steps a child takes toward adult knowledge, so in this sense we do know the process by which a child develops. However, we do not know the *experiences* that are crucial, and how the child uses them to change the grammar in the adult direction. In addition, there are many questions remaining for clinical practice.

As an example, suppose it is discovered that the child must master a crucial building block of some domain before he or she can handle more complex forms. Would intense exposure to that simpler form then be the best way to advance the development? Suppose that exposure to the complex form, however rare, is what is necessary to assign the simple form its correct structure. Then any amount of drilling with the simple form alone could in fact fail to help. Such a case might be double *wh*-questions, such as "Who bought what?" We have searched corpora of parent-to-child speech (CHILDES)[2] and found these to be extremely rare in the input. Is it therefore safe to ignore them? These questions reveal most clearly the property of *wh*-questions as "variables"; that is, as demanding a set answer. To a double *wh*-question, there is no alternative except to give the paired sets: "Martin bought eggs, Aisha bought butter, James bought bread." Elaborate practice at answering *wh*-questions of a simpler sort, which require the child to name one object, might teach nothing about this property. One can imagine other pitfalls of simplification; for example, teaching vocabulary by always ensuring that the item being taught could be pointed at, and so reducing uncertainty. How then would the child ever learn to use context to guess a meaning?

Training studies are a relatively untapped source of important information that would be of immediate utility to clinicians facing these questions. A few such studies have been done, but there is much room for further work in which success at particular structures is the target of inquiry, and different strategies of presenting the components are tried. In our thinking about such matters, we have been influenced by the work of Mary Wilson,[3] who designs computer software for language training of children with language disabilities. She has argued that the limited amount of time typically devoted to specialized language therapy cannot possibly suffice to remediate profound language learning difficulties. After all, if the normal language environment has not

sufficed in the 5 years or so of the child's life to date, why would 1 or 2 hours a week more of general language exposure do much good? Of course, there are clinicians who argue that the whole world of the child must be enriched by making increased opportunities for language connections in the classroom and at home. However, such interventions are often framed in very general terms, and the question of what a particular child needs may be overlooked. With luck, somewhere in that rich language surround, the right piece of information, the right insight, or the right encouragement will make its mark. This is a reasonable approach given our uncertainty about the right pieces and how to deliver them, but it is hardly adequate and runs the risk of missing crucial ingredients. Linguistics must be brought to bear on this larger question of how to effect language growth, or these advances in assessment will not fulfill their promise.

CONCLUSION

The results of our studies have revealed that some important and deep properties of language show up as causing difficulty in children with language disorders, regardless of dialect. We have not yet shown how dialects work, how a disorder might be manifest within the special properties of dialect, or whether young dialect speakers with disorders might show difficulties in learning to code-switch. Cooperation among linguists, dialect experts, and speech therapists is necessary to carry the questions forward. Joint expertise must also be brought to bear on the larger question of how to effect language growth, or these advances in assessment will not fulfill their promise.

REFERENCES

1. Pearson BZ. Theoretical and empirical research for dialect-neutral language assessment: contributions from theoretical and applied linguistics to communication disorders. Semin Speech Lang 2004; 25:xxx–xxx
2. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk, 3rd ed. Mahwah, NJ: Lauwrence Erlbaum; 2000
3. Wilson M. Sequential Software for Language Intervention and Development. Winooski, VT: Laureate Learning Systems; 1996