LSHSS

Clinical Forum

Alternative Assessment of Language and Literacy in Culturally and Linguistically Diverse Populations

Sandra P. Laing
University of Alabama, Tuscaloosa
Alan Kamhi
University of Oregon, Eugene

ABSTRACT: Over- and underdiagnosis of language and literacy problems are common with low-socioeconomic status ethnically and racially diverse children. In recent years, a number of alternative assessment procedures have been developed that reduce some of the biases inherent in norm-referenced standardized tests. Problems and recent solutions to the use of norm-referenced testing will be discussed, with a focus on processing-dependent and dynamic assessment procedures.

KEY WORDS: assessment, multicultural, language, literacy

The population in the United States has gradually become more ethnically and racially diverse. Future projections suggest that this trend will continue at a stable rate through the year 2050 (U.S. Bureau of the Census, 2000). The more diverse the population, the more likely it is that clinicians will encounter families from cultural, ethnic, and linguistic backgrounds that differ from their own (Hanson, 1992). These changing demographics demand a rapid solution to the paucity of culturally valid and reliable assessment instruments for culturally and linguistically diverse (CLD) populations.

A number of authors (Battle, 2002; Franklin, 1992; Hanson, 1992) have provided some general guidelines for how clinicians can improve their cultural sensitivity to better provide unbiased and accurate assessment and diagnosis of language abilities and language disorders in CLD children. Culturally sensitive clinicians have an awareness of different cultures as well as specific experiences interacting with individuals from different cultures. They also have knowledge of customs, beliefs, and values of different cultures, as well as knowledge of the specific language differences that characterize the dialects or languages CLD children speak.

It is well accepted that the accurate assessment of spoken and written language skills in CLD children should not depend solely on the use of standardized norm-referenced test procedures (Taylor & Payne, 1983; Washington, 1996; Wyatt, 1995). As many clinicians know, test taking is a cultural phenomenon that by its very nature is biased against children who are raised in families that do not provide many out-of-context test-like situations (e.g., Battle, 2002). In the first part of this article, some problems and recent solutions to the use of norm-referenced testing will be discussed. In the remainder of the article, two promising procedures to provide unbiased assessment of CLD children's language and literacy will be presented—the use of processing-dependent measures and dynamic assessment measures.

PROBLEMS WITH NORM-REFERENCED TESTING

The problems with norm-referenced measures have been well documented (Brice, 2002; Washington & Craig, 1992;

0161–1461/03/3401–0044

Wilson, Wilson, & Coleman, 2000). Perhaps the most familiar of these problems are content bias, linguistic bias, and disproportionate representation in normative samples.

Content Bias

Content bias occurs when test stimuli, methods, or procedures reflect the assumption that all children have been exposed to the same concepts and vocabulary or have had similar life experiences. For example, test stimuli are typically derived from the concepts and vocabulary used in White middle-class school settings (Washington, 1996) and on familiar interaction patterns in mainstream culture (Boykin, 1977; van Kleeck, 1994; Wyatt, 1995). Children from CLD backgrounds may perform more poorly on standardized measures relative to mainstream children because of variations in life experience, socialization practices, and early literacy experiences (Stockman, 2000). To illustrate, labeling or pointing to objects or actions may not be a typical language experience for Hispanic children (Anderson, 2002; Peña, Quinn, & Iglesias, 1992). Thus, errors on the Peabody Picture Vocabulary Test–III (PPVT–III; Dunn & Dunn, 1997) may reflect a deficit in receptive vocabulary, a lack of familiarity with the interaction pattern of pointing to pictures, a lack of familiarity with the vocabulary of English, or a combination of these factors.

Linguistic Bias

Linguistic bias may also be associated with the use of standardized tests and refers to a disparity between (1) the language or dialect used by the examiner, (2) the language or dialect used by the child, and (3) the language or dialect that is expected in the child's responses. For example, if the examiner is using Standard American English (SAE) and the child is using African American Vernacular English (AAVE), the child may perform differently than expected not because of a language disability, but because of a mismatch between the child's dialect and the dialect of the testing tool. Poor performance reflects a linguistic difference and not impairment.

Attempts to make adjustments to standardized tests in order to reduce linguistic bias have led to both over- and underidentification of impairments. Overidentification has been the most prevalent problem, attributing "errors" to dialect differences. Underidentification can occur when an examiner assumes that a child who belongs to a particular racial/ethnic group will use the dialect associated with that group (Wilson et al., 2000). Thus, differences in performance may be attributed to dialect when in fact they are errors.

Researchers have attempted to reduce over-/under-identification by omitting culturally suspect items or not counting the items wrong (Cole & Taylor, 1990). For example, for an AAVE speaker, test items containing the copula or auxiliary *be* could be omitted from final scoring, or the child could be given credit for the items. Although there are clearly advantages to eliminating dialectally sensitive items from standardized tests, there is some concern that it may be difficult to determine whether an incorrect response is due to cultural/linguistic differences or a language disorder.

Disproportionate Representation in Normative Samples

Historically, standardized tests have not included CLD populations in their normative samples. Recently, test developers have endeavored to solve this problem by including more representative proportions of diverse populations in the normative groups. The most recent normative samples for the Test of Language Development–Primary–3 (TOLD–P–3; Newcomer & Hammill, 1997) and the Test of Language Development–Intermediate–2 (TOLD–I–2; Newcomer & Hammill, 1988) include 15% African Americans and 8% Hispanics. The Test of Adolescent and Adult Language–3 (TOAL–3; Hammill, Brown, Larsen, & Wiederholt, 1994) includes African American (15%) and Hispanic children (8%) in its broad normative sample. Similarly, 15% of the normative group for the Test of Language Development–Intermediate–3 (TOLD–I–3; Hammill & Newcomer, 1997) are African American and 10% are Hispanic.

Including children from diverse backgrounds in the normative sample does not necessarily eliminate the problem of over- or underidentifying CLD children with language disorders. Consider, for example, the 1981 version of the Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981), which contained a normative sample that was 14% African American. More than 90% of 105 typically developing low-income African American preschool children performed below the mean (Washington & Craig, 1992). The most recent version of this measure, the PPVT–III (Dunn & Dunn, 1997), includes more African American children than the PPVT–R (34%) and more closely resembles the distribution of the population in the United States (Stockman, 2000). At-risk preschool African American children (N = 59) who were administered the PPVT–III performed within normal limits, albeit at the low normal range (M = 91; Washington & Craig, 1999). Adjusting the normative sample to represent CLD populations accurately may do nothing more than decrease the mean distribution of the normative sample to the point where CLD children still perform below the mean, but within normal age limits. In other words, they are not performing like their age-matched peers, so what was once a language disorder (e.g., on the PPVT–R) is now a language difference on the PPVT–III.

Content bias, linguistic bias, and disproportionate representation in normative samples have been and continue to be problems that researchers and clinicians will address in the assessment of CLD populations. Perhaps the only certain way to eliminate bias is to develop an assessment instrument that is designed for specific CLD groups. Seymour, Roeper, and de Villiers (in preparation) are in the process of developing a norm-referenced test for African American children, and Peña, Gutierrez-Clellen, and Iglesias (personal communication, May 21, 2002) are developing a norm-referenced test for Hispanic children

who are learning English as a second language. These tests should be able to differentiate CLD children with language differences accurately from those who have language disorders. They may not, however, show whether CLD children have language skills comparable to those of mainstream children.

CRITERION-REFERENCED MEASURES

Even the best norm-referenced test is no substitute for other measures of assessment. For example, the use of criterion-referenced measures is usually much better than the use of norm-referenced tests for determining whether a CLD child has a language disorder and in providing clinical focus (Battle, 2002). A criterion-referenced measure compares a child's performance on a specific skill, grammatical structure, or linguistic concept to independently predetermined criteria. Such measures make it possible to consider the social context in which communication occurs and how language is used by the culture. Criterion-referenced measures can be designed by the clinician using language, materials, contexts, and interaction patterns that are familiar to the child, thus eliminating the content and linguistic bias that is often associated with norm-referenced measures.

Unfortunately, criterion-referenced measures are only as good as the developmental data on which they are based. Because of the lack of well-established developmental information on certain CLD populations, it is sometimes difficult to set valid criteria for mastery of specific linguistic forms. Some researchers (e.g., Terrell & Terrell, 1993) have recommended that comparisons be made to the language patterns of parents or caregivers. For example, the same tasks can be given to the parent and child so that patterns of dialectical variation can be differentiated from patterns of errors (Terrell, Arensberg, & Rosa, 1992). Responses that do not match the parents' would be considered disorders. On the other hand, it may be inaccurate to assume that the linguistic behaviors of a child emulate those of an adult (Washington, 1996). To illustrate, a number of studies suggest that the frequency of use of AAVE forms differs between children and adults (Ramer & Rees, 1973; Ratusnik & Koenigsknecht, 1976). Given these limitations, it is essential to supplement the use of criterion-referenced measures with other alternative assessment procedures.

ALTERNATIVE ASSESSMENT PROCEDURES

Norm- and criterion-referenced measures are the most common ways to assess children's language and literacy. There are, however, a number of other procedures used to evaluate language. These include language sampling procedures, ethnographic interviewing techniques, processing-dependent measures, and the use of dynamic assessment. Language sampling and ethnographic interviewing should be a part of any evaluation for a child from a CLD population (Battle, 2002; Brice & Montgomery, 1996; Cheng, 1991; Crago & Cole, 1991; Kayser & Restrepo, 1995; Mattes & Omark, 1991; Westby, 1990). Language samples and ethnographic interviewing are beneficial in the diagnosis of CLD populations because they can be obtained in natural settings using language, communication partners, and interaction patterns that are familiar to the family and child. Examination of language and literacy under these conditions should yield reliable information about linguistic competency in CLD children. The remainder of the article will highlight two additional procedures, processing-dependent techniques and dynamic assessment techniques, and their prospective roles in the assessment of language and literacy in CLD populations.

Processing-Dependent Measures

Campbell, Dollaghan, Needlemen, and Janosky (1997) speculated that it might be possible to reduce testing bias for CLD children by using measures that placed more emphasis on processing abilities and less emphasis on prior language knowledge and experience. A processing-dependent task is minimally dependent on prior knowledge or experience. Examples of processing-dependent tasks include various memory tasks (e.g., digit span, working memory, nonword repetition), certain perceptual tasks (e.g., discrimination of rapidly presented tones, sequencing tones presented in rapid sequence), and competing stimuli tasks (e.g., filtered words, auditory figure ground, competing words). In their study, Campbell et al. administered three processing-dependent and one knowledge-dependent measure to 156 typically developing children between the ages of 11 and 14 (M = 12;6 [years;months]). Two thirds of the children were African American. The processing-dependent measures were a nonword repetition task (NWR-T; Campbell, Needleman, Riess, & Tobin, 1995), The Competing Language Processing Task (CLPT; Gaulin & Campbell, 1994), and The Revised Token Test (RTT; McNeil & Prescott, 1978). The knowledge-dependent measure was the Oral Language Scale (OLS) from the Woodcock Language Proficiency Battery–Revised (Woodcock, 1991).

The NWR-T consisted of 24 nonsense words varying in length from two to four syllables that bore no resemblance to familiar English words (see Table 1). To minimize the influence of articulation skill on performance, none of the words contained consonant clusters, later developing phonemes, or lax vowels. Scoring is based on the total number of phonemes repeated correctly. Phoneme substitutions and omissions are counted as errors; phoneme additions and distortions are considered correct.

The CLPT is a working memory task. The CLPT requires children to respond to the truth value of a series of statements and then recall words from each statement. For example, children were asked to judge the truth value of sentences (e.g., "Trees have leaves" and "Trains can fly") and then recall the last word in the sentence. Performance is based on the number of words recalled correctly.

The RTT is designed to measure children's ability to carry out increasingly complex spoken commands (e.g., Touch the blue circle; Touch the big, blue circle; Touch the big blue circle and the small red square). This is similar to

Table 1. Nonword repetition task.[a]

Instructions: "Now I will say some made-up words. Say them after me exactly the way that I say them."

1.	ʃəɑmədæk	13.	bəʃɑrɪs
2.	bakum	14.	fəmɪp
3.	bɪməwʊkəs	15.	bɑn
4.	ʃəmɑs	16.	gɛfʊsəpəm
5.	mæfot	17.	məb
6.	bɪθəsə|	18.	kosəpʊm
7.	puk	19.	kɪv
8.	funəθ	20.	gədæ|əb
9.	pɛfəmɛk	21.	fɑ|dɛs
10.	səvitəs	22.	bɪponəfɪd
11.	muk	23.	sə|
12.	θɪgə|ɛkən	24.	pɛθətɑnɪs

[a] From "Reducing Bias in Language Assessment: Processing-Dependent Measures," by T. Campbell, C. Dollaghan, H. Needleman, & J. Janosky, 1997, *Journal of Speech, Language, and Hearing Research, 40,* p. 525. Copyright 1997 by the American Speech-Language-Hearing Association. Reprinted with permission.

the Concepts and Directions subtest of the Clinical Evaluation of Language Fundamentals–3 (CELF–3; Semel, Wiig, & Secord, 1995).

The OLS is designed to measure spoken language ability and consists of three subtests (Oral Vocabulary, Listening Comprehension, and Verbal Analogies). Each subtest yields a standard score and is combined to yield an oral language score composite.

The CLD children obtained significantly lower scores (M = 91.19, SD = 13.22) than the mainstream children (M = 107.84, SD = 15.38) on the knowledge-based task (OLS), but did not differ on any of the processing-dependent measures (NWR-T, CLPT, RTT). The findings in this study can be regarded as preliminary evidence that the processing-dependent measures were less biased toward typically developing CLD children than was the knowledge-dependent measure. More importantly, these findings led researchers to ask whether similar processing-dependent measures would be useful in the identification of impairment in CLD populations.

To answer this question, Rodekohr and Haynes (2001) administered the NWR-T and CLPT tasks from Campbell et al. (1997) to 40 African American and Caucasian children between the ages of 7;0 and 7;3 (M = 7;2). The knowledge-dependent measure was the Test of Language Development–Intermediate–2 (TOLD–I–2; Newcomer & Hammill, 1988). The children were divided into four groups: Group 1: 10 African American children with normal language who spoke African American English (AAE), Group 2: 10 African American children with language impairment who spoke AAE, Group 3: 10 Caucasian children with normal language who spoke Southern English (SE), Group 4: 10 Caucasian children with language impairment who spoke SE.

All three measures differentiated the children with language impairments from those with normal language independent of race or dialect. More importantly, performance on the knowledge-based and processing-dependent measures was consistent with Campbell et al.'s (1997) findings: African American children performed comparably to the Caucasian children on the processing-dependent measure, but significantly poorer than the Caucasian children on the knowledge-based measure. These findings support the notion that African American children often exhibit knowledge-based deficiencies on norm-referenced language tests despite having processing abilities comparable to those of mainstream children.

Given the findings in these two studies, the use of processing-dependent measures appears to be helpful in identifying CLD children who have normal processing abilities but some language-based knowledge deficiencies. In contrast, processing-dependent measures appear to provide no advantage in differentiating children with language impairments, other than the diagnostic value in identifying both processing and knowledge deficiencies in these children. There is one concern regarding the use of NWR-Ts with CLD children. Although the stimuli on the task developed by Campbell et al. (1995) were designed not to resemble English words or syllables, there may be some items that bear a resemblance to real words in other languages. In these instances, stimuli might have to be modified for use with children whose first language is not English.

A working memory task like the CLPT may provide a practical alternative to the NWR-T because it is not affected by phonological factors. In an unpublished study, Laing (2002) examined the performance of 73 third- and fourth-grade children on a verbal working memory measure. Participants were 24 typically developing African American children, 33 African American children with language disorders, and 16 typically developing Caucasian children. Children were administered a working memory task (Appendix A). The task consisted of nine sentence pairs: three sets of two statements, three sets of three statements, and three sets of four statements. Children were required to listen to every sentence in each set and then indicate whether the statement was true or false. A chime that was sounded after the final statement in each set signaled the children to recall the last word of each sentence in the set. Items that included cities, states, and landmarks were modified to be familiar to children growing up in Memphis, Tennessee (see items 9, 15, 20, 23, Appendix A). As with the CLPT, performance was based on the number of words accurately recalled.

Performance on the verbal working memory task was similar for the two groups of typically developing children. Table 2 shows the grade levels, means, and standard deviations for the three groups of children. A one-way analysis of variance (ANOVA) revealed a significant difference between groups, $F(2, 70) = 13.31$, $p = .0001$. Typically developing children performed better than the African American children with language impairment on the working memory task; however, post hoc testing revealed that only the difference between typically developing African American children and African American children with language impairment was significant. Effect size testing with Cohen's d (Cohen, 1988) suggested that both typically developing groups were significantly different from the group of African American children with language

Table 2. Mean *(M)* and standard deviation *(SD)* on Gottardo, Stanovich, and Siegel's (1996) verbal working memory task for third- and fourth-grade African American and Caucasian children.

	African American		*Caucasian*
	Typical (n = *24*)	*Language disorder* (n = *33*)	*Typical* (n = *16*)
M	18.8	15.4	18.4
SD	3.33	2.34	2.28

*$p < .01$

impairment. Therefore, these results are similar to previous findings regarding the NWR-T, the CLPT, and the RTT: The typically developing Caucasian and African American children performed comparably on the verbal working memory task.

Processing-dependent measures are easy to administer and score, making them useful, convenient techniques for clinicians. Although existing research suggests that they may be most useful in identifying processing deficiencies that may or may not coexist with other impairments, their use is a promising avenue for future investigations.

Dynamic Assessment

Another promising assessment alternative in CLD populations is the use of dynamic assessments. The concept of dynamic assessment stems from the work of Vygotsky (1978) and his conceptualization of a *zone of proximal development*. The zone of proximal development is the difference between a child's current level of independent performance on a task and how he or she succeeds at that same task given guided assistance. With dynamic assessment, one determines not only current level of functioning, but also the best means to facilitate further learning. Dynamic assessment is sometimes characterized as diagnostic teaching because of its focus on improving performance.

Test–teach–retest. Perhaps the most familiar approach to dynamic assessment is to test, teach, and then retest. This method of dynamic assessment has been shown to differentiate stronger and weaker language learners in Puerto Rican, African American, and Native American preschool and kindergarten children (Lidz & Peña, 1996; Ukrainetz, Harpell, Walsh, & Coyle, 2000). For example, Lidz and Peña (1996) compared the test–teach–retest method of dynamic assessment to a standardized measure of vocabulary in predicting how well two bilingual preschool children responded to training and how their responsiveness ratings related to posttest scores on the standardized measure. Both children performed well below the mean (standard score of 55) on the Expressive One-Word Picture Vocabulary Test–Revised (EOWPVT–R; Gardner, 1990) before two, 20-min teaching sessions that targeted naming objects and pictures. Prompts and cues were provided based on the child's responses to instruction and were scored using a Likert scale according to the level of responsiveness, examiner effort in training, and transfer to new contexts.

The two children responded very differently to the intervention. One child demonstrated high responsivity, needed minimal examiner effort, and showed a high degree of learning transfer. The other child evidenced only slight responsivity, required maximum support from the examiner, and showed only a moderate degree of transfer. Consistent with this pattern of modifiability, the first child increased his score on the EOWPVT–R by two standard deviations, whereas the second child showed little change. This case study shows that standardized measures of vocabulary provided little information about learning potential.

In a subsequent study, Gutierrez-Clellen and Peña (2001) demonstrated how a dynamic assessment of vocabulary knowledge might be used. Two Latin American, bilingual preschool children participated in two, 30-min mediated learning experiences (MLEs). The MLE sessions were designed to increase the children's ability to label objects in their environment. The two children performed similarly on the EOWPVT–R (Child A, SS = 67; Child B, SS = 71) before the MLE sessions, but differed in their performance on the Comprehension subtest of the Stanford-Binet Intelligence Scale (78/100; Thorndike, Hagen, & Sattler, 1986) and the Preschool Language Scale–3 (Child A, 4/10 correct; Child B, 7/10 correct; Zimmerman, Steiner, & Pond, 1992). The Learning Strategies Checklist (Peña, 1993) was used to rate children's (a) attention, (b) comparative behavior, (c) planning, (d) self-regulation, (e) transfer, and (f) motivation. An overall rating of modifiability was obtained using a Likert scale for rating examiner effort, child responsivity, and transfer or generalization of information taught to the posttest session.

The 2 children received different modifiability ratings, which were reflected on their performance on the EOWPVT–R after mediation. Child A was judged as moderately responsive to the MLE and demonstrated no change in performance on the EOWPVT–R after training (SS: 67 pre; 67 post). In contrast, Child B was judged to be highly modifiable and made gains in performance on the EOWPVT–R (SS: 71 pre; 86 post). These findings indicate that it may be more appropriate to diagnose Child A with a language learning problem than Child B. Responsiveness to the MLE would appear to be an excellent way to distinguish CLD children with language differences from those with language disorders.

Task/stimulus variability. Another method of dynamic assessment is to modify the way tests are presented. Recall that one of the problems with standardized tests is that they do not accurately reflect differences in life experiences that may differ from mainstream culture. These differences may translate into poorer performance by CLD children in standardized testing situations. For instance, research has suggested that African American children's academic performance is higher in classrooms that provide children opportunities to learn through social and physical interaction, including the use of cooperative learning, movement, and music (Collins & Tamarkin, 1982; Dillon, 1989; Hale-Benson, 1986). Assessments that are presented in naturalistic environments and are action–object oriented may

provide more appropriate measures of performance for African American children raised in high-energy homes (Franklin, 1992).

This was the dynamic assessment approach taken by Fagundes, Haynes, Haak, and Moran (1998). In a preliminary study, Haynes, Haak, Moran, Rice, and Johnson (1995) investigated the performance of 70 African American and Caucasian children attending Head Start programs on the Preschool Language Assessment Instrument (PLAI; Blank, Rose, & Berlin, 1978). All of the children scored below middle-class northern children (the normative group), but the African American children scored significantly below the Caucasian children. Given previous research findings that suggested that African American children may perform better on assessments that are action–object oriented and presented in naturalistic environments, Fagundes et al. questioned whether administering the PLAI in the context of thematic activities would improve African American children's performance.

In the more recent study (Fagundes, Haynes, Haak, & Moran, 1998), the PLAI stimulus items were administered to coincide with story time, snack time, and arts and crafts activities that children were engaged in. For example, a question on the PLAI is "If this bowl (point to picture of bowl) were filled all the way up with play dough, could we pour these (point to picture of marbles) inside?" The same question was administered in the Preschool Language Assessment Instrument–Thematic version, (PLAI–T; Blank, Rose, & Berlin, 1978) in the context of manipulation and play with the actual objects (i.e., bowl, play dough, and marbles). The test was administered with and without these activities to 12 low-socioeconomic status (SES) African American children and 12 low-SES Caucasian children. The African American children performed comparably to their age-matched peers when the stimulus items were administered in the context of thematic activities. As in the previous study, African American children performed significantly worse than their age-matched peers when the test was administered in its standardized format. Thus, embedding language forms in realistic thematic contexts appears to be a useful way to differentiate language disorders from language differences in CLD children. Children whose performance does not improve when tasks are administered in the context of ongoing situations are most at risk for linguistic and/or academic difficulty through more traditional intervention.

Other commonly used language norm-referenced tests, such as the CELF–3 (Semel et al., 1995) or the TOLD-I–3 (Hammill & Newcomer, 1997), could also be modified using task/stimulus variability techniques to make them less static and decontextualized. For example, the items in the Concepts and Directions and Word Class subtests of the CELF–3 could be administered with object or picture support in a more thematic context. To illustrate, rather than use the black line drawings as stimuli for the Concepts and Directions subtest, three-dimensional shapes might be used in the framework of a Simon Says activity (e.g., "Simon says point to the circle at the beginning of the line"). Similarly, word classes might be administered using real objects or pictured stimuli and asking children to group like objects together. The child would be presented with a button, a small doll's shirt, and a small doll's chair. The clinician would label each of the objects and then ask the child to select the two that go together. Pictures could be used to present the stimuli that are not easily represented with objects (e.g., eagle, wing, window). For the items that cannot easily be portrayed with objects or pictures, a more interesting thematic context might be used. For example, for the sentence recall task, children can pretend they are robots and use a computer voice to repeat the sentences. Children whose performance does not improve in contextualized formats may be at risk for linguistic and academic problems that may be addressed through intervention (Paul, 2001).

Graduated prompting. Another type of dynamic assessment is the use of graduated prompting techniques. Determining whether a child is stimulable for the production of sounds not currently being produced is an example of this type of dynamic assessment. How well a child responds to graduated prompts can help to determine which language forms and structures to target and the amount of improvement a child might be expected to make in intervention (Bain & Olswang, 1995). With graduated prompting, assessment and intervention phases for a particular behavior occur simultaneously.

Bain and Olswang (1995) questioned whether a graduated prompting procedure would be a better predictor of response to treatment than a number of standardized measures of language. Standardized measures included the Sequenced Inventory of Communication Development–Revised (SICD–R; Hedrick, Prather, & Tobin, 1984), the PPVT–R (Dunn & Dunn, 1981), and mean length of utterance (MLU) in morphemes. The dynamic assessment procedure was a graduated prompting technique that provided children with varying levels of supportive verbal cues to elicit a variety of communicative intentions. The supportive cues ranged from a general statement (least supportive) to specific elicitation questions, sentence completion statements, indirect models, and direct models (most supportive). The least supportive cue was provided first. If a child failed to respond to this cue, a more explicit cue was provided. Responses were scored based on the amount of support required. A score of 6 indicated that the child needed little support (general statement); a score of 1 indicated that the child needed a lot of support (direct model).

In the study, 15 children with specific expressive language impairments (age 30–36 months) were provided with individual language treatment three times a week for 40-min sessions. Treatment targeted increased use of communicative intentions and increases in utterance length. Assessments took place 3 weeks before and after the intervention sessions. The graduated prompting measure was found to be the best predictor of changes in MLU (r = .73, $p < .001$). The only standardized measure that was significantly correlated with language progress was the receptive–expressive language age discrepancy score based on performance on the SICD (r = .49, $p < .05$).

Graduated prompting techniques have also been used to predict early reading abilities (Spector, 1992). Laing, Kamhi, and Catts (1997) examined how well static and

dynamic measures of phonological awareness predicted early reading performance. Participants in their study were 72 typically developing kindergarten and first-grade children who were administered a dynamic and static segmentation measure in the fall and spring, as well as measures of reading performance (Woodcock Reading Mastery Test–Revised [WRMT–R]; Woodcock, 1997). The static segmentation task was the Words to Sounds subtest of the Sawyer's Test of Awareness of Language Segments (Sawyer, 1987). The dynamic segmentation measure was developed by Catts (Laing et al., 1997) and consisted of a series of graduated prompts (see Appendix B). When a child was unable to respond correctly on an item, prompts were given to provide increasing support to the child in completing the task. Scoring was based on the number of prompts required to succeed on the task. For example, the least supportive prompt included an exact repetition of the instructions, "Listen while I say the word very slowly. Now can you say each of the sounds in the word?" The most supportive prompt included having the examiner model segmentation of the sounds in the word while physically assisting the child in using blocks to represent each sound.

Table 3 presents the means, standard deviations, and *t*-test statistics for the fall and spring administrations of the static and dynamic segmentation tasks for kindergarten and first-grade children. As can be seen in this table, all of the children improved significantly on their performance on all of the phonological awareness tasks (static and dynamic) from the fall to the spring. However, as can be seen in Table 4, the static segmentation task administered in the fall was not significantly related to either of the spring reading measures for kindergarten or first grade. In contrast, the dynamic measure of segmentation administered in the fall significantly predicted word attack skills for the kindergarten children ($\rho = .60$, $p < .01$) and word identification skills for the first-grade children ($\rho = .60$, $p < .01$). Thus, the dynamic measure of phonological awareness predicted reading better in the spring than did the static measure of phonological awareness. This study provides some initial evidence that a dynamic measure of phonological awareness may be a better predictor of reading than static measures of phonological awareness.

Dynamic assessment procedures that include graduated prompting techniques can be used to evaluate proficiency in specific skills and behaviors, predict how well a child may perform when given direct instruction, and determine which instructional techniques are most effective in promoting learning. After identifying the skills that are essential to complete a particular task, a series of prompts ranging from least to most supportive are constructed. A final step in the development of a graduated prompting measure is to devise a scoring system so that progress can be charted as children acquire the skill. Scoring should depend on the number of prompts incorporated and be logically ordered. Here is how the Word Order subtest from the TOLD–I–3 might be modified.

Clinician: I'm going to say some words that you must make into a sentence. Use all the words and put them together to make a sentence: girl, sad, is, the

Client: girl, sad, is, the

Clinician: (Prompt 1) Listen one more time. I'm going

Table 3. Mean *(M),* standard deviation *(SD),* and *t* tests for static and dynamic phonological awareness measures for kindergarten and first grade.

Group	*Fall static*	*Spring static*	t	*Fall dynamic*	*Spring dynamic*	t
Kindergarten	(n = 19)	(n = 19)		(n = 18)	(n = 18)	
M	.68	3.79	3.71*	16.67	45.7	7.39*
SD		1.4	5.3		21.05	37.72
First Grade	(n = 17)	(n = 17)		(n = 18)	(n = 18)	
M	5.82	12.88	26.6*	68.18	105.47	6.78*
SD	5.4	4.3		32.6	9.99	

*$p < .01$

Table 4. Correlation coefficients for fall phonological awareness measures and spring reading measures.

	Word Identification		*Word Attack*	
Measure	*Kindergarten*	*First grade*	*Kindergarten*	*First grade*
Static segmentation	.01	.15	.33	.04
Dynamic segmentation	.11	.60*	.60*	.35

*$p < .01$

to say some words that you must make into a sentence. Use all the words and put them together to make a sentence: girl, sad, is, the

Client: girl, sad, is, the

Clinician: (Prompt 2) I want you to use all the words and put them together to make a sentence that makes sense. I'll give you a hint, the first word in the sentence is "The."

Client: The, girl, sad, is

Clinician: (Prompt 3) Listen to each of the sentences I am going to read to you and tell me whether it is ok or silly. The is to brick, I like cake, She orange up scream etc. (Paul, 2001, p. 526).

Scoring would simply follow the number of prompts required. A score of *3* might be given for success after prompt 1, a score of *2* after prompt 2, and a score of *1* after prompt 3. The same set of prompts could be used in teaching sentence combining using simple sentences.

Slight modifications in these prompts would be necessary for the Grammatical Comprehension and Malapropism subtests. Placing heavy intonational stress on the target words to highlight the words that students should be attending to might alter prompt 2. Prompt 3 could include accompanying pictures.

For the Picture Vocabulary subtest, prompt 2 could be a brief discussion about physical attributes and/or functions about related pictures. Prompt 3 might entail having children generate two-word descriptions for pictures after which the clinician would point to the appropriate picture.

These examples illustrate how norm-referenced language tests can be modified to include graduated prompts. Researchers and clinicians who find that children perform well in response to graduated prompting techniques in dynamic assessment sessions should take steps to ensure that these skills generalize to less structured, more natural situations in which children must function socially and academically.

CONCLUSION

Over- and underdiagnosis of language and learning disabilities in CLD children will continue to be a problem because the reference point for identifying a disorder is arbitrary. If one uses psychometric criteria, norm-referenced tests will be used to identify language and learning disabilities. Children who fall below a certain percentile of performance will be identified as language or learning disabled. If one renorms existing tests to better capture the characteristics of the CLD group or develops instruments that are designed for specific groups, fewer children will be inaccurately identified as disabled.

Until those measures are available, researchers and clinicians must pursue other avenues for accurate diagnosis and intervention for CLD populations such as the ones discussed in this article. The use of processing-dependent and/or dynamic assessment measures with CLD populations is appealing for a number of reasons. They are not biased toward life experience, socialization practices, or literacy knowledge, and they are quick and easy to administer. Further, many nonword repetition and working memory measures are included in currently existing standardized measures. Clinicians can measure working memory using NWR-Ts or using standardized digit span and word span tasks (Montgomery, 2002). Performance on nonword repetition and working memory measures has been found to be highly correlated with language impairment and second-language vocabulary acquisition in adults and children (Atkins & Baddeley, 1998; Bishop, North, & Donlan, 1996; Gottardo, Stanovich, & Siegel, 1996; Papagno, Valentine, & Baddeley, 1991; Service & Kohonen, 1995). In other words, when children perform poorly on processing-dependent measures, there is a high likelihood that they will have some type of language learning difficulty. Future research should continue to explore the use of these measures to help identify CLD children with language disorders.

REFERENCES

Anderson, R. (2002). Practical assessment strategies with Hispanic students. In A. E. Brice (Ed.), *The Hispanic child* (pp. 143–184). Boston: Allyn & Bacon.

Atkins, P., & Baddeley, A. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19,* 537–552.

Bain, B., & Olswang, L. (1995). Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. *American Journal of Speech-Language Pathology, 4*(1), 81–91.

Battle, D. (2002). Language development and disorders in culturally and linguistically diverse children. In D. Bernstein & E. Tiegerman-Farber (Eds.), *Language and communication disorders in children* (pp. 354–386). Boston: Allyn & Bacon.

Bishop, D., North, T., & Donlan, C. (1996). Nonword repetition as a behavioral marker for inherited LI: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 36,* 1–13.

Blank, M., Rose, S., & Berlin, L. (1978). *Preschool Language Assessment Instrument.* New York: Psychological.

Boykin, A. (1977). Experimental psychology from a Black perspective: Issues and examples. In W. Cross (Ed.), *Final report from the Third Conference on Empirical Research in Black Psychology* (pp. 53–77). Washington, DC: National Institute of Education.

Brice, A., & Montgomery, J. (1996). Adolescent pragmatic skills: A comparison of Latino students in ESL and speech-language programs. *Language, Speech, and Hearing Services in Schools, 27,* 68–81.

Brice, A. E. (2002). *The Hispanic child.* Boston: Allyn & Bacon.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40,* 519–525.

Campbell, T., Needleman, H., Riess, J., & Tobin, M. (1995, June). *Bone lead level and language processing performance.* Poster session presented at the annual meeting of the Symposium on Research in Child Language Disorders, Madison, WI.

Catts, H. (1997). The early identification of language-based reading disabilities. *Language, Speech, and Hearing Services in Schools, 28,* 86–89.

Cheng, L. (1991). *Assessing Asian language performance.* Oceanside, CA: Academic Communication Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, P., & Taylor, O. (1990). Performance of working-class African American children on three tests of articulation. *Language, Speech, and Hearing Services in Schools, 21,* 171–176.

Collins, M., & Tamarkin, C. (1982). *Marva Collins' way.* Los Angeles: Jeremy P. Tarcher.

Crago, M., & Cole, E. (1991). Using ethnography to bring children's communicative and cultural words into focus. In T. M. Gallagher (Ed.), *Pragmatics of language: Clinical practice issues* (pp. 99–132). San Diego, CA: Singular.

Dillon, D. (1989). Showing them that I want them to learn and that I care about who they are: A microethnography of the social organization of a secondary low-track English reading classroom. *American Educational Research Journal, 26,* 227–259.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test–Revised.* Circle Pines, MN: American Guidance Service.

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

Fagundes, D., Haynes, W., Haak, N., & Moran, M. (1998). Task variability effects on the language test performance of southern lower socioeconomic class African American and Caucasian five year olds. *Language, Speech, and Hearing Services in Schools, 29,* 148–157.

Franklin, M. (1992). Culturally sensitive instructional practices for African American learners with disabilities. *Exceptional Children, 59,* 115–122.

Gardner, M. (1990). *Expressive One-Word Picture Vocabulary Test–Revised.* Novato, CA: Academic Therapy Publications.

Gaulin, C., & Campbell, T. (1994). A procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79,* 55–64.

Gottardo, A., Stanovich, K., & Siegel, L. (1996). The relationship between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third grade children. *Journal of Experimental Psychology, 63,* 563–582.

Gutierrez-Clellen, V., & Peña, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools, 32,* 212–224.

Hale-Benson, J. (1986). *Black children: Their roots, culture and learning styles.* Baltimore: The Johns Hopkins University Press.

Hammill, D., Brown, V., Larsen, S., & Wiederholt, J. (1994). *Test of Adolescent and Adult Language–3.* Austin, TX: Pro-Ed.

Hammill, D., & Newcomer, P. (1997). *Test of Language Development–Intermediate–3.* Austin, TX: Pro-Ed.

Hanson, M. (1992). Ethnic, cultural, and language diversity in intervention settings. In E. Lynch & M. Hanson (Eds.), *Developing cross-cultural competence: A guide for working with young children and their families* (pp. 1–18). Baltimore: Brookes.

Haynes, W., Haak, N., Moran, M., Rice, R., & Johnson, V. (1995, November). *The Preschool Language Assessment Instrument: Performance differences in rural southern African American and White headstart children.* Paper presented at the Convention of the American-Speech-Language-Hearing Association, Orlando, FL.

Hedrick, D., Prather, E., & Tobin, A. (1984). *Sequenced Inventory of Communication Development–Revised.* Seattle, WA: University of Washington Press.

Kayser, H., & Restrepo, M. (1995). Language samples: Elicitation and analysis. In H. Kayser (Ed.), *Bilingual speech-language pathology: An Hispanic focus* (pp. 265–286). San Diego, CA: Singular.

Laing, S. (2002). [Performance on Gottardo, Stanovich & Siegel's (1996) verbal working memory task for 3rd and 4th grade typically developing and language impaired African American and Caucasian children]. Unpublished raw data.

Laing, S., Kamhi, A., & Catts, H. (1997; November). *Dynamic assessment of phonological awareness in school age children.* Paper presented at the Annual Convention of the American-Speech-Language-Hearing Association, Boston.

Lidz, C., & Peña, E. (1996). Dynamic assessment: The model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *Language, Speech, and Hearing Services in Schools, 27,* 367–384.

Mattes, L., & Omark, D. (1991). *Speech and language assessment for the bilingual handicapped.* Oceanside, CA: Academic Communication Associates.

McNeil, M., & Prescott, T. (1978). *The Revised Token Test.* Baltimore: University Park Press.

Montgomery, J. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11,* 77–91.

Newcomer, P., & Hammill, D. (1988). *Test of Language Development–Intermediate–2.* Austin, TX: Pro-Ed.

Newcomer, P., & Hammill, D. (1997). *Test of Language Development–Primary–3.* Austin, TX: Pro-Ed.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign language vocabulary learning. *Journal of Memory and Language, 30,* 331–347.

Paul, R. (2001). *Language disorders from infancy through adolescence: Assessment and intervention.* St. Louis, MO: Mosby-Yearbook.

Peña, E. (1993). *Dynamic assessment: A non-biased approach for assessing the language of young children.* Unpublished doctoral dissertation, Temple University, Philadelphia, PA.

Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A non-biased procedure. *Journal of Special Education, 27,* 269–280.

Ramer, A., & Rees, N. (1973). Selected aspects of the development of English morphology in Black American children of low socioeconomic background. *Journal of Speech and Hearing Research, 16,* 569–577.

Ratusnik, D., & Koenigsknecht, R. (1976). Influence of age on Black preschoolers' nonstandard performance of certain phonological and grammatical forms. *Perceptual and Motor Skills, 42,* 199–206.

Rodekohr, R., & Haynes, W. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders, 34,* 1–18.

Sawyer, D. (1987). *Test of Awareness of Language Segments.* Austin, TX: Pro-Ed.

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals–3.* San Antonio, TX: Psychological.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16,* 155–172.

Seymour, H., Roeper, T., & de Villiers, J. (in preparation). *Dialect Sensitive Language Screener.* San Antonio, TX: Psychological.

Spector, J. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353–363.

Stockman, I. (2000). The new Peabody Picture Vocabulary Test–III: An illusion of unbiased assessment. *Language, Speech, and Hearing Services in Schools, 31,* 340–353.

Taylor, O., & Payne, K. (1983). Culturally valid testing: A proactive approach. *Topics in Language Disorders, 3,* 8–20.

Terrell, S., Arensberg, K., & Rosa, M. (1992). Parent–child comparative analysis: A criterion-referenced method for the nondiscriminatory assessment of a child who spoke a relatively uncommon dialect of English. *Language, Speech, and Hearing Services in Schools, 23,* 34–42.

Terrell, S., & Terrell, F. (1993). African-American cultures. In D. Battle (Ed.), *Communication disorders in multicultural populations* (pp. 3–37). Boston: Andover.

Thorndike, R., Hagen, E., & Sattler, J. (1986). *Stanford-Binet Intelligence Scale: Fourth edition.* Riverside, CA: Riverside.

Ukrainetz, T., Harpell, S., Walsh, C., & Coyle, C. (2000). A preliminary investigation of dynamic assessment with Native American kindergartners. *Language, Speech, and Hearing Services in Schools, 31,* 142–154.

U.S. Bureau of the Census. (2000) *Population estimates program.* Washington, DC: Author.

van Kleeck, A. (1994). Potential cultural bias in training parents as conversational partners with their children who have delays in language development. *American Journal of Speech-Language Pathology, 3*(1), 67–78.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Washington, J. (1996). Issues in assessing the language abilities of African American children. In A. Kamhi, K. Pollock, & J. Harris (Eds.), *Communication development and disorders in African American children: Research, assessment, and intervention* (pp. 35–54). Baltimore: Brookes.

Washington, J., & Craig, H. (1992). Articulation test performances of low-income, African-American preschoolers with communication impairments. *Language, Speech, and Hearing Services in Schools, 23,* 203–207.

Washington, J., & Craig, H. (1999). Performance of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test–III. *Language, Speech, and Hearing Services in Schools, 30,* 75–82.

Westby, C. (1990). Ethnographic interviewing. *Journal of Childhood Communication Disorders, 13,* 110–118.

Wilson, W., Wilson, J., & Coleman, T. (2000). Culturally appropriate assessment: Issues and strategies. In T. Coleman (Ed.), *Clinical management of communication disorders in culturally diverse children* (pp. 101–128). Boston: Allyn & Bacon.

Woodcock, R. (1991). *Woodcock Language Proficiency Battery–Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. (1997). *Woodcock Reading Mastery Test–Revised.* Circle Pines, MN: American Guidance Service..

Wyatt, T. (1995). Language development in African American English child speech. *Linguistics and Education, 7,* 7–22.

Zimmerman, I., Steiner, V., & Pond, R. (1992). *Preschool Language Scale–3.* San Antonio, TX: Psychological.

Received June 3, 2002
Accepted October 6, 2002
DOI:10.1044/0161-1461(2003/005)

Contact author: Sandra P. Laing, PhD, The University of Alabama, Speech and Hearing Center, Box 870242, Tuscaloosa, AL 35487-0242. E-mail: slaing@cd.as.ua.edu

APPENDIX A. VERBAL WORKING MEMORY TASK[a]

We are going to play a memory game. It has two parts. You are going to hear some statements and I want you to answer TRUE or FALSE (yes or no for Grade 1 and 2 or children who cannot tell you what true and false mean). Then you will hear a chime and I want you to give me the last word of each of the sentences in the group just given. It does not matter if you give the words in the same order as the sentences.

e.g., Set 1: You are a man/woman/boy/girl. (Ask so that answer is TRUE)
I am sitting/standing (Ask so that answer is FALSE) Ok. Now recall.
Set 2: This room is purple.
It is raining.

1. The sun rises in the evening.
2. Trees lose their leaves in spring.
(evening, spring)
3. Cars have four wheels.
4. Cows and pigs eat meat.
5. A red traffic light means "Go."
(wheels, meat, go)
6. We get milk from cows.
7. Plants need water and light to grow.
8. In winter it is warm
9. The pyramid is in Memphis.
(cows, grow, warm, Memphis)
10. We read from right to left.
11. Lettuce and peas are vegetables.
(left, vegetables)
12. Centimeters are used for measuring.
13. Elephants have gray spots.
14. Some birds have fur.
(measuring, spots, fur)
15. Arkansas is close to Tennessee.
16. A motorcycle can move faster than a bicycle.
17. An apple is a fruit.
18. Fish swim in the sky.
(Tennessee, bicycle, fruit, sky)
19. People can buy groceries in stores.
20. Little Rock is the capital of Arkansas.
(stores, Arkansas)
21. We use a thermometer to tell time.
22. Boiling water is hot.
23. Memphis is on the shore of the Ms.
(time, hot, Mississippi)
24. A football is round.
25. We sleep at night.
26. Insects have eight legs.
27. A feather is heavier than a rock.
(round, night, legs, rock)

Give starred items only to subjects who get one of the four-item sets correct.

*28. Some birds fly north in winter.
*29. The earth travels around the sun.
*30. Purple, red, and big are colors.
*31. The United States is the smallest country in the world.
*32. Tadpoles become frogs.
(winter, sun, colors, world, frogs)

[a] From "The Relationship Between Phonological Sensitivity, Syntactic Processing, and Verbal Working Memory in the Reading Performance of Third Grade Children," by A. Gottardo, K. Stanovich, & L. Siegel, 1996, *Journal of Experimental Psychology, 63,* p. 582. Copyright 1996 by Elsevier Science. Adapted with permission.

APPENDIX B. DYNAMIC ASSESSMENT OF PHONOLOGICAL AWARENESS (SEGMENTATION)[a]

Instructions: Say to the child, "We are going to play a word game. I will say a word, and I want you to say each of the sounds in the word. For example, if I say *shave,* you would say /sh/, /e/, /v/." Proceed to test items, providing prompts when necessary.

Test Items

1. say
2. me
3. zoo
4. of
5. fish
6. chew
7. leaf
8. sick
9. feet
10. map
11. top
12. big
13. goat
14. flew
15. smoke
16. bleed
17. ghost
18. box
19. slaps
20. twist

1. "Listen while I say the word very slowly." Say the word slowly for the child. "Now can you say each of the sounds in the word?" Give score of 5 if correct response after prompt #1. If the child is still unsuccessful, go to prompt #2.

2. "What's the first sound you hear in _____?" If the child says the first sound correctly, say "Now can you tell me each of the sounds?" If the first sound is incorrect or no response is given, say, "Try to tell me just a little bit of the word." Give a score of 4 if correct response after prompt #2. If the child does not produce the first sound, skip prompt #3 and #4 and go to prompt #5.

3. If the child correctly produces the first sound but not the others, say "_____ is the first sound in _____. What sound comes next?" "And next?" Give a score of 3 if correct response after prompt #3. If the child is still unsuccessful, go to prompt #4.

4. "There are _____ sounds in _____." "What are they?" Give a score of 2 if correct response after prompt #4. If incorrect, go to prompt #5.

5. "Watch me." Now model segmentation of the word and place a token in a square as you say each of the sounds, then repeat the word as a whole. Say to the child "Try to do what I did." Give a score of 1 if correct response after prompt #5. If incorrect, go to prompt #6.

6. "Let's try together." Now model segmentation of word with the child and assist in moving the blocks. "Try to do it yourself." Repeat once if necessary. No points are given if an incorrect response is given or if correctly given after prompt #6.

Give a score of 6 for each item 1–20 that is responded to correctly with no prompt.

[a] Developed by Hugh Catts. Reprinted with permission.