

Dynamic Assessment of Diverse Children: A Tutorial

Vera F. Gutiérrez-Clellen

San Diego State University, CA

Elizabeth Peña

University of Texas at Austin

Traditional language assessment focuses on the identification of language impairment by determining the child's current level of performance in a given area relative to the performance of his or her peers. However, a child's limited test performance may reflect different learning experiences or a lack of educational opportunity, and not necessarily language deficits. Children from culturally and linguistically diverse (CLD) backgrounds may exhibit depressed test performance, yet their performance may not reflect their true abilities or learning potential. On the other hand, CLD children with language impairment may be at risk for under-referral if language difficulties are believed to be language differences. For these children, clinicians must be able to use appropriate methods to differentiate children with a language difference from those with a language disorder. Developing a new test, local norms, or good

translations for a test may not reveal children's language-learning skills because these approaches focus only on the child's current level of performance rather than on the child's ability to respond to learning experiences (Peña, Quinn, & Iglesias, 1992).

Dynamic assessment (DA) represents an alternative or supplemental approach to traditional language assessments with CLD children. However, there is great variability across DA methods and applications and little information concerning which approaches may be best suited to differentiate language differences from disorders. The purpose of this article is twofold: (a) to consider available DA methods and procedures as they apply to the differentiation of language difference from disorder, and (b) to propose an assessment protocol that may be useful to minimize misdiagnosis.

ABSTRACT: Dynamic assessment (DA) has been advocated as an alternative and/or supplemental approach to traditional standardized testing with children who are culturally and linguistically diverse (CLD). However, there is great variability across DA methods and applications, as well as limited information concerning which methods and procedures may be best suited to differentiate language disorder from difference. In this tutorial, DA methods are compared with respect to their assessment applications. Next, an assessment protocol recommended for the identification of language disorder versus difference is described. Finally, examples of two Spanish-English bilingual children are used to show how the protocol may be useful to assess children's language-learning potential and to minimize misdiagnosis.

KEY WORDS: dynamic assessment, children, cultural diversity

THE MODEL

DA is based on Vygotsky's model of cognitive development (Vygotsky, 1986). Within this model, the child's knowledge develops in social interactions with more capable others. These experiences are culturally mediated and gradually become internalized as higher cognitive functions. According to Vygotsky (1978), learning takes place in the "zone of proximal development" (ZPD). A child acquiring new information initially requires maximal assistance from an adult, but eventually is able to assume greater responsibility for the activity as the information becomes internalized. The ZPD lies between the level of performance the child can reach unassisted, and the level that can be attained when adult assistance is provided. In DA, the goal is to determine the "size" of the ZPD. That is, the goal is to establish the amount of change that can be induced during interactions with the examiner during the assessment process.

Clinically, DA can assist clinicians in determining when and how to intervene. For example, children who show limited change during the assessment may require extensive assistance in order to facilitate changes in language behavior. These children would be candidates for specific intervention in the areas revealed by the assessment. In contrast, children who show significant changes during the assessment, and who can maintain those changes, may not need language services. Their initial poor performance might suggest unfamiliarity with the test situation or culturally based language differences. The information gathered during the DA can allow the examiner to make inferences about underlying learning processes and the nature of intervention needed to effect change. However, some DA methods focus more on determining appropriate interventions, whereas others focus on the goal of helping children perform at their optimum level. Thus, not all DA methods are equally useful for making diagnostic decisions with CLD children.

DA METHODS AND APPLICATIONS

Historical Background

Clinical and school psychologists first considered DA methods during the 1920s (for a historical review, see Lidz, 1987). At that time, psychologists interested in examining an individual's capacity for improvement began to question intelligence tests as measures of learning *ability*. During the 1950s, studies compared IQ gains of low scoring students using coaching (e.g., review and discussion of errors similar to those on the IQ test, explanation of instructions, and practice with similar problems). Yet, these approaches had significant limitations (for a critique, see Anastasi, 1981). Little information was available about the coaching procedures, and gains had limited generalization to other cognitive measures. During this time, there also were attempts at testing the *limits*. That is, the examiner determined the highest level at which a child could respond, given probing. For example, Jedrysek, Klapper, Pope, and Wortis (1972) used a series of probes to assess the level at which a child could succeed. Although the approach was not dynamic, at that time, it represented an alternative approach to traditional testing.

In the 1960s, psychologists began to apply a test-teach-retest model to cognitive assessment. In Israel, Feuerstein (1979, 1980) developed the Learning Potential Assessment Device (now called the Learning Propensity Assessment Device), which he used to determine and modify the low functioning of Moroccan Jewish immigrant children. For these children, Feuerstein argued that there was a discontinuity of cultural transmission related to their past histories of immigration and displacement. As children moved from their rural communities to Casablanca and then to Israel, their socialization into the culture of their ancestors was disrupted, and as a result, they had limited opportunity to learn how to think about problems, such as those used in cognitive testing. The DA provided needed mediation and

showed that these children had the potential to learn when appropriate learning opportunities were provided.

The application of DA to cognitive assessments expanded during the 1970s and 1980s. Numerous tools for assessing children's learning potential were developed and different procedures were explored (e.g., Budoff, 1974; Lidz, 1991, 1996). An important advance in the development of DA approaches was the assessment of transfer, or generalization of learning, to other tasks (e.g., Campione & Brown, 1987). The application of DA to cognitive assessments was expanded to include academic achievement assessments, such as reading and mathematics. Given that a detailed discussion of these studies is beyond the scope of the present article, the next sections will focus on the methods and applications of DA to language testing.

DA Methods

Methods for the DA of language can be categorized as (a) testing the limits, (b) graduated prompting, and (c) test-teach-retest. Although these three methods share general features, testing the limits and graduated prompting are more appropriate to determine readiness for progress in intervention. In contrast, test-teach-retest methods are better suited to differentiate disorders from differences.

Testing the limits. Carlson and Wiedl (1978, 1992) compared and validated testing the limits in a series of studies with young children. Within their testing the limits approach (which they later called dynamic assessment), traditional test procedures were modified by providing elaborated feedback or verbalization. Carlson and Wiedl differentiated between simple feedback and elaborated feedback. In giving simple feedback, the examiner simply indicated whether the answer was correct. In contrast, elaborated feedback included feedback about the correctness of the child's answers, an elaboration of the reasons why the answer was correct or not, and an explanation of the principles involved in the task. For verbalization, children were asked to describe the test question and then to verbalize how they arrived at an answer. Carlson and Wiedl found that the use of elaborated feedback and verbalization was more effective with 5- to 10-year-olds from various ethnic groups than standard assessments or assessments that included simple feedback. Their studies using cognitive and academic measures with children from diverse developmental and ethnic backgrounds indicated that modifications of the test conditions through feedback led to higher estimates of ability on posttest measures than traditional test conditions, and resulted in a reduction in test anxiety (Carlson & Wiedl, 1978, 1992; Ginsburg, 1986).

Applying this model to vocabulary testing, Peña (n.d.) found that providing feedback during testing improved the performance of CLD children on the Expressive One-Word Picture Vocabulary Test (EOWPVT, Gardner, 1979) in comparison to the performance of those children who did not receive feedback during testing. During administration of the test, examiners explained to children why each response was correct or incorrect, and if incorrect, they asked the child to try again. For example, when children were correct on a given item, the examiner said, "Good,

that was a special name.” On the other hand, when children were incorrect, the examiner pointed it out by acknowledging the child’s response and how it could be plausible: “Yes, you do eat that. Can you think of the special name?” The type of information that is gained from testing the limits includes (a) whether the child understands the task and (b) whether the child has competence in a given area that is not revealed by testing.

Another way of testing the limits is by modifying the administration of a test to incorporate a clinical interview condition. This approach focuses on generating questions to help children understand how they are thinking about a test question, and through this process, facilitate their awareness of targeted skills. Ginsburg (1997) demonstrated that children from CLD backgrounds were able to demonstrate their knowledge better in a clinical interview condition than on a traditional test focused on the correctness of their responses.

Peña (2001) (see also Ginsburg, 1997) incorporated this technique when working with CLD children and found it useful to gain an understanding of how children view the assessment task from their point of view, much like an ethnographic approach. In contrast, with the Carlson and Wiedl testing the limits approach, the examiner provides no feedback about the “correctness” of children’s responses. Instead, children are asked questions such as, “How did you know that?” and “What would happen if?” in an attempt to understand how they are thinking about a question and why they responded the way they did. Through this questioning, children often change their responses, demonstrating that they know the answer to the test questions. However, this approach requires that children have sufficient metalinguistic and metacognitive skills. If these conditions are met, testing the limits through a clinical interview may help children perform at higher levels and, possibly, reduce bias in assessment.

Clinicians need to be cautioned that results obtained through these techniques, although having high face validity (i.e., they may appear to assess children’s understanding of the task and their true competence in a target area), may have questionable reliability (i.e., they may not yield consistent results across testers or examinees). In order to ensure some level of reliability, clinicians should plan the feedback or verbalization before the testing and be consistent in the use of the techniques. To this date, the effectiveness of the testing the limits approaches for differentiating language disorders from differences has not yet been evaluated empirically. Thus, this approach should not be used without additional DA procedures.

Graduated prompting. Graduated prompting procedures attempt to identify the ZPD by providing the child with a hierarchy of predetermined prompts. In general, the child’s readiness to learn specific targets is determined by the level that the child attains given minimal and maximal prompts. Prompts are designed to vary in the level of contextual support they provide. The child’s responses to the test procedures and new tasks are used to draw predictions about the child’s response to intervention.

Campione and Brown (1987) determined a child’s modifiability based on the number of prompts needed to elicit a desired response and the level of transfer of learned

skills to novel tasks using cognitive measures. Based on these procedures, differences in learning ability are expected in the transfer of skills to new items. Transfer distance was conceptualized as a continuum from no transfer to near transfer, far transfer, or very far transfer. For example, children who are taught with a set of toys that objects and people have specific names (as opposed to ambiguous labels) demonstrate no transfer if they continue using “this thing” or “that” to refer to a different set of toys representing the same objects. The same children would demonstrate near transfer if they use specific labels to refer to these objects when they are presented in photographs. Far transfer would be demonstrated if the children use specific reference in other contexts, such as in play or in the classroom. Finally, very far transfer would be demonstrated if the children use specific labels to name pictures in a decontextualized test situation (e.g., when answering comprehension questions).

Applications of Campione and Brown’s prompting procedure suggest that the number of prompts needed to elicit targets and the transfer distance to novel items can predict children’s gain scores posttesting. The use of graduated prompting procedures has been applied extensively in language assessment with very young children (Bain & Olswang, 1995; Olswang & Bain, 1996). For example, Olswang, Bain, and Johnson (1992) designed a DA protocol using a hierarchy of verbal cues to determine whether children at the one-word stage of development were ready to produce two-word utterances. Children were taught single words using six types of cues, for example, modeling (“This is a baby”), modeling with an elicitation question (“This is a baby. What is it?”), modeling with an object obstacle (withholding the object until the child attempted to produce the word), and so forth. Although children had similar pre-intervention profiles, they differed on how well they responded to the cues and in the types of cues needed to elicit two-word utterances. This indicated differences in their responsiveness to the cues, and, ultimately, it suggested differences in their language readiness.

Graduated prompting also has been applied to the assessment of children’s phonemic awareness (Evans, Maschmeyer, & McFarlane, 1996; Spector, 1992). For example, children who had difficulty segmenting a word received a fixed set of graduated cues. These cues were (a) pronouncing the target word slowly, (b) asking the child to identify the first sound in the word, (c) cueing the child with the number of sounds in the word, (d) modeling segmentation using pennies placed in squares to represent the number of sounds in the word, and (e) working hand over hand with the child while pronouncing the segments. Each response then was scored based on the level of assistance needed. For example, a correct response without prompt received a maximum score of 6, a correct response after the first prompt in the hierarchy received a score of 5, and so forth. Spector (1992) demonstrated that the procedure was a better predictor of progress in learning to read (as measured by word recognition scores at the end of the school year) than static measures such as standard assessments, in which children are tested without an examiner’s prompting or help. The results corroborate previous

advantages reported in studies with young language learners (Olswang & Bain, 1996) that the graduated prompting procedure had significant advantages over static measures.

Although graduated prompting may be helpful for making predictions about responsiveness to intervention, the procedure may not be effective for all children. Long and Olswang (1996) noted that there might not be a direct relationship between children's responsiveness to the graduated prompting procedure and future learning growth curves. That is, there may be children who appear responsive during the DA, but who show limited growth over time. In addition, available studies are based on very young children who are producing one-word utterances (e.g., Olswang & Bain, 1996), or children with IQs as low as 60 (Campioni & Brown, 1987). Thus, the effectiveness of this method with older children or children with higher cognitive skills is unclear.

Furthermore, task hierarchies that focus on discrete skill learning with CLD children may not facilitate transfer to other tasks because they tend to be highly structured or decontextualized (Norris & Hoffman, 1990). Some researchers argue that children with language impairments have difficulty learning under natural conditions, and therefore, may benefit from approaches that are more structured (e.g., Fey, 1986). Other researchers argue that approaches for CLD children that focus on discrete decontextualized task hierarchies do not allow them to reveal their true learning potential (Reyes, 1992). Thus far, there is no evidence that graduated prompting may be useful to differentiate language disorders from differences. Until more research becomes available, this method should be limited to examining a child's response to intervention.

Test-teach-retest. Researchers examining DA with CLD populations have long advocated for the use of a test-teach-retest paradigm to assess children's learning potential. Most of the work using this paradigm has originated in the field of cognitive and educational psychology. For example, Budoff (1974) first proposed it as a means to equalize differences in students' experiences that affect their performance on cognitive tests. Within this paradigm, the examiner first identifies deficient or emerging skills that may be related to a lack of mediated learning experiences (MLEs) with that skill. Then, the examiner provides an intervention (the MLE) designed to modify the child's level of functioning in the targeted areas. By teaching the principles of the task, the test situation changes from an *evaluative* interaction (typical of traditional test situations) to a *teaching* interaction where the examiner maximizes the child's feelings of competence. The performance on the posttest (retest phase) serves as an indicator of the child's modifiability following training.

The test-teach-retest method has been used with children with disabilities as well as with children from diverse ethnic groups. Some of these applications are highly individualized and non-standardized (Feuerstein, 1979; Lidz & Thomas, 1987), whereas other applications are applied in standard fashion (e.g., Budoff, 1987). Highly standardized approaches maximize reliability, but may threaten validity. On the other hand, non-standardized, individualized

approaches have high face validity, but reliability may be more difficult to establish (Peña, 1996).

Budoff (1987) applied the test-teach-retest method with children from various developmental and ethnic groups using a battery of tests or "instruments" from the Learning Potential Assessment Device that involved cognitive tasks such as block designs, completion of patterns, and so forth. The teaching tasks consisted of coaching procedures that provided increasing support for the child. For each test, training procedures, such as the use of verbal explanation, prompts, examples, or models, were specified. By incorporating a teaching component in the assessment process, Budoff's test-teach-retest paradigm offered children opportunities for learning and demonstrating their ability to learn as a measure of their true competence. As will be seen in the examples to be presented shortly, the approach can help clinicians differentiate language differences from disorders. However, Budoff's original teaching phase included relatively standardized training procedures, which may not meet the individual needs of some children (Lidz, 1991).

In contrast, Tzuriel and Klein (1987) proposed using somewhat unstructured intervention procedures with relatively standardized goals. For example, for assessing the modifiability of children's analogical thinking, they developed the Children's Analogical Thinking Modifiability Instrument (CATM). Using this, the examiner takes 18 blocks varying in color, shape, and size and asks questions such as: "Small blue square [is to] small blue circle [as] large yellow square [is to]...?" After pretesting, the examiner reviews the dimensions of the materials (i.e., size, color, shape, picture names) and the rules of problem solution (what children need to think of in order to solve the problem). These goals address general aspects of performance, which include understanding analogical principles and using a consistent approach to search for the missing block. However, the examiner can provide any type of mediation depending on the child's needs. Posttesting is done using alternative test forms of the CATM. Tzuriel and Klein (1985) applied this method to the cognitive assessment of preschool children from mainstream and low socioeconomic status backgrounds, as well as that of children with mental retardation and children in special education. Their research indicated that children's modifiability could be differentiated on the basis of posttest scores.

Most DA studies that focused on differentiating language disorders from culturally based differences have used the test-teach-retest paradigm. However, unlike Budoff's approach, greater individualization of the teaching component in DA has been proposed in order to address socialization differences across children. For example, Gutiérrez-Clellen, Peña, and Quinn (1995) showed that there was significant variability among Head Start teachers in the types of support provided for narrative learning, perhaps because teachers assume that children come to school with specific narrative experiences. Yet, because children from diverse backgrounds may vary in their narrative socialization experiences, they may not display expected narrative skills in classroom contexts, and as a result, their narrative differences may be confused with disorders. The DA of narratives, therefore, should include an ethnographic

examination of the types of narrative experiences available to the child, the types of interactions in those narrative learning situations, and teacher mediation strategies (Gutiérrez-Clellen et al., 1995).

Within the test-teach-retest model, the MLE proposed was developed based on these observations and interviews with the teachers, and it focused on individual children's needs. Children who had narrative difficulties were thought to be able to answer questions about story facts (e.g., what, where, etc.) but not able to relate the narrative events cohesively or to explain the actions of the characters. The MLE focused on organizing the story into basic components (e.g., setting, events, consequences, and reactions) using questioning following each narrative episode. The questions were not focused on the story content, but on directing children's attention to narrative structure and the process of reconstructing literature narratives. For example, the examiner would ask the children to think about the important features of each narrative episode and would discuss how the episodes interconnected in a cohesive whole. The examiner also would show children how to think about story comprehension questions. Posttesting consisted of retelling a different narrative followed by questions about the critical features of each episode and the relationships among the episodes. Modifiability measures included evidence of the use of mediated strategies for summarizing information and for monitoring comprehension of events, as well as gains in the production of correct answers to story comprehension questions.

Miller, Gillam, and Peña (2001) also applied a test-teach-retest in their program, "Improving Children's Narrative Abilities." Within this application, children tell a story based on a wordless book, which is evaluated for the use of story components, episode structure, ideas, and language. Two MLEs then are used to target two areas in which the particular child shows difficulty. The DA includes test-retest gains as well as ratings of listener effort and student responsiveness as measures of modifiability. When appropriate mediated narrative experiences are provided, it is predicted that children who may not have had access to these experiences prior to the assessment would show high modifiability. That is, children with different narrative experiences but who have normal narrative learning ability would show significant changes on the posttest. Because the intervention phase of the DA is not standardized, the procedures should result in greater opportunity for change. In contrast, children with true language disorders are expected to show limited modifiability as well as limited generalization to other narrative tasks. Their language needs would require additional interventions and increased effort.

Measures of Modifiability

In addition to selecting DA procedures with relatively unstructured mediation procedures (as opposed to predetermined probe hierarchies), clinicians should select appropriate measures of change carefully. Budoff's (1987) original approach examined only posttest gains as a measure of change. Yet, gain scores may not provide sufficient

information about a child's modifiability. DA studies found that modifiability scales can be reliable and useful to evaluate change in DA (e.g., Gutiérrez-Clellen, Brown, Robinson-Zañartu, & Conboy, 1998; Peña et al., 1992).

Modifiability scales using Likert-type ratings (adapted from Lidz, 1991, and Peña, 1993) were used to examine behaviors, such as the child's ability to attend, self-regulate, or use the adult as a resource. Ratings also were used to assess the child's overall responsiveness to mediation, the child's ability to transfer the new skill to a novel task, and the intensity of effort required by the mediator to induce change. For example, attention was rated on a scale from 1 to 5 (1 = high distractibility and minimal attention; 2 = prompting/repetition needed more than 50% of the time; 3 = prompting/repetition needed less than 50% of the time; 4 = attentional prompts needed only at the beginning of the mediation session; and 5 = orienting response and on-task behavior without prompts). Based on these criteria, Gutiérrez-Clellen and colleagues were able to obtain a Cronbach's reliability of .90. Peña (2000) found that the modifiability scale scores classified children with low and typical language ability with moderate to high diagnostic accuracy. That is, the scales were accurate at identifying a child as having a language disorder (diagnostic sensitivity) or normal language (diagnostic specificity).

Peña and colleagues (Peña, 1993, 1996, 2000; Peña, Iglesias, & Lidz, 2001; Peña et al., 1992) investigated children's ZPD by incorporating an MLE during the assessment process. Change was assessed using both posttest gains and modifiability scales. Group comparisons indicated that, although children from diverse backgrounds with typical and low language ability were not differentiated on the basis of a single-word vocabulary test at the pretest, the two ability groups performed differently on posttest scores after mediation. The groups also exhibited different modifiability scores, which were measured using ratings of responsiveness, examiner effort, and transfer obtained immediately after each MLE session. The combination of posttest scores and modifiability ratings appeared to be more useful clinically to differentiate differences from disorders than pretest scores alone.

Research evaluating DA procedures for establishing the true language needs of diverse children also suggests an analysis of their responses as a supplemental measure of change. Some children may not exhibit change in the *number* of correct responses during posttesting, but in the *quality* of the responses given to the test. Thus, for some children, the changes may be qualitative rather than quantitative. Gutiérrez-Clellen et al. (1998) assessed bilingual students with a history of learning disabilities using antonym and synonym tasks. The MLE experience was focused on teaching the principles underlying the production of synonyms (but not antonyms). Modifiability was determined by language gains on the mediated task, evidence of transfer based on post-MLE scores on antonyms, qualitative changes in children's responses from pre- to post-MLE, and behavioral observations using modifiability scales. The scales appeared accurate at predicting limited language-learning behavior and modifiability in children who made significant or no language

gains post-MLE. They were less accurate at assessing the potential for language change in children who exhibited some language gains. The study illustrated the application of gain scores, modifiability scales, and an analysis of children's responses as important modifiability measures with this population.

SUMMARY

The test-teach-retest method can assist clinicians in differentiating a language disorder from a difference by incorporating specific learning experiences during the assessment process and then evaluating children's responsiveness to these interactions. These learning experiences may not have been available to many diverse children due to differences in educational experiences. The alternative procedures just discussed appear to be the best approach to determine whether diverse children have specific language performance difficulties that are not easily modifiable during the DA and that would require further interventions.

DA methods such as testing the limits or graduated prompting have not been evaluated for their effectiveness in differentiating language disorders from differences in CLD children and have several limitations. In contrast, the test-teach-retest approach appears to be best suited for the identification of language disorders because the examiner makes no assumptions about children's ability based on previous knowledge and experiences with test situations and tasks. The approach should be applied using relatively unstructured mediation activities that explicitly teach the principles of the task (i.e., what is expected), and using different probes (i.e., test items should not be used for teaching). A structured approach using a hierarchy of prompts may not elicit expected responsiveness in CLD children because these approaches do not teach directly how to solve problems or how to respond to specific language-testing questions or prompts. Measures of change should include posttest scores, modifiability ratings, and qualitative analysis of children's responses as important indicators of the child's ability to change. In the next sections, a DA protocol is illustrated that (a) incorporates intervention in the assessment process in order to assess the child's responsiveness to the examiner's mediation, (b) examines the child's ability to change using a variety of measures, and (c) determines the factors that facilitate change for specific language areas in order to establish appropriate educational or intervention goals.

CASE STUDIES

Recent research has shown that vocabulary tests are not accurate identifiers of language impairment (Gray, Plante, Vance, & Henrichsen, 1999). This research indicates that these tests have unacceptable diagnostic sensitivity or specificity for the identification of language impairment in children from majority culture groups. For the Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R,

Gardner, 1990), diagnostic sensitivity and specificity was reported to be only 71%. These rates suggest that almost 30% of children likely would be misdiagnosed using this measure. Similarly, vocabulary tests applied to CLD children do not differentiate children with and without language disorders using either national or local norms (Peña & Quinn, 1997). When national norms were applied to African American and Puerto Rican children with and without language impairment, approximately 91% of the children were identified as having a vocabulary deficit. On the other hand, when local norms were applied, none of the children suspected of having a vocabulary deficit was identified (Peña et al., 2001). The assessment approach was not appropriate to identify children with lexical acquisition deficits.

Gray and her colleagues indicated that word-learning measures may identify lexical acquisition skills of children with language impairments with better accuracy than performance on vocabulary tests. Indeed, in a DA condition, CLD children with language differences but not disorders significantly increased their posttest scores after the examiner provided appropriate mediation, and also showed higher modifiability during the mediated learning experiences (Peña et al., 1992; Peña et al., 2001). For many children, low scores on a vocabulary test may not mean vocabulary deficits, but culturally based differences. It is when children, whether or not from CLD backgrounds, obtain scores below expectations that DA is critically needed. Some of these children may score low as a result of individual differences or other factors, and some may have true vocabulary learning deficits.

To our knowledge, there is no information concerning the extent to which children who are not CLD and who can perform at expectation levels would show high posttest scores within DA. However, research in the field of school psychology found that whereas CLD and non-CLD children scored significantly different on the pretest of cognitive measures, elaborated feedback had a compensatory effect on the posttest performance, significantly reducing differences between the two groups (Carlson & Wiedl, 1980). Although the non-CLD children demonstrated posttest gains, their improvements were not as great as those of the CLD children because of the fact that their initial scores were already high.

When vocabulary tests are used, clinicians typically draw inferences about children's vocabulary learning ability. Yet, these inferences are made without knowledge about how children would have performed given specific mediation opportunities. In contrast, within a test-teach-retest model, the goal is to use a given language measure as a tool to assess language-learning skills. The approach does not use vocabulary scores to determine the presence of vocabulary deficits. Instead, the examiner uses the posttest scores and observations of the child's modifiability to estimate vocabulary learning in a particular area. Although the following examples illustrate the use of DA to examine children's vocabulary learning, the approach can be used to examine other language domains using other tests of language or criterion-based measures (Gutiérrez-Clellen & Quinn, 1993; Gutiérrez-Clellen et al., 1995; Gutiérrez-Clellen et al., 1998). As it will be illustrated next, DA

examines the child's modifiability by focusing on how the child's performance changes in a given area, such as vocabulary tasks, when appropriate mediation experiences are provided.

The examples were drawn from a larger DA study of preschool African American and Latin American children that was reported in Peña et al. (2001). In this study, 55 children participated in MLEs that focused on single-word labeling. These children were compared to 22 children who were tested but did not receive MLEs. The two case studies to be described next are based on two children in the MLE group. They were selected to show how to evaluate different modifiability profiles in order to make diagnostic decisions.

Participants

The two children (Child A and Child B) were Latin American, bilingual (Spanish-English) speakers, living in the same neighborhood in Philadelphia, Pennsylvania. They were enrolled in a Head Start program and attended the same class. The children's language abilities were determined based on classroom observation of peer interaction, teacher report, and parent report. In addition, the children were tested using three language tools: (a) the EOWPVT-R (Gardner, 1990) and its Spanish translation to obtain an estimate of the child's vocabulary; (b) the Comprehension subtest of the Stanford-Binet Intelligence Scale (CSSB, Thorndike, Hagen, & Sattler, 1986) to assess how well the children responded to questions requiring functions, descriptions, and explanations; and (c) 10 items (5 expressive and 5 receptive) of the Preschool Language Scale (PLS, Zimmerman, Steiner, & Pond, 1978) to assess knowledge of school-based language concepts. The latter provides a raw score from 1 to 10 for each child. The children profiled here were selected because, although they scored similarly on the EOWPVT-R pre-MLE, they showed different posttest profiles and different levels of modifiability.

Child A. Child A was a female, aged 4:0 (years:months), who spoke mainly Spanish but used some English words in interaction with her teacher and with English-speaking peers. Observation of peer interaction during free play indicated that she responded mainly in single words to peer initiations and used single words plus gestures to communicate with English-speaking peers. Her teacher was concerned about her interaction with other children and her very limited language skills.

Standardized testing was conducted mainly in Spanish, allowing Child A to respond in English, if necessary. Scores reported are conceptual scores, which give credit to knowledge of an item regardless of the language used. Compared to the national norms, her EOWPVT-R standard score was 67 ($M = 100$, $SD = 15$). On the CSSB, her standard score was 78 ($M = 100$, $SD = 16$). Finally, on the PLS (adapted version, PLS-A), she was able to respond correctly to 4 out of 10 possible items.

Based on this profile, it was unclear whether Child A had specific vocabulary difficulties because many CLD children may have difficulty with vocabulary tests and academic concepts as a result of different or limited

educational experiences. Child A's limited participation in school activities also could be related to differences in exposure to specific educational experiences. Her test scores on the EOWPVT-R and the CSSB were obtained as part of a larger effort to compare the effectiveness of using local norms for various language measures. Yet, comparisons with local norms obtained at the same Head Start program were not very useful either. She scored 67 on the EOWPVT-R, which was within the average of 72.76 ($SD = 8.08$) based on local norms (Peña et al., 2001). Her score of 78 on the CSSB was 1.26 SD below her Head Start peers ($M = 94.05$; $SD = 12.71$). For this child, performance on these measures was not consistent across tasks and communicative interactions. Thus, the goal of the DA was to determine the potential role of specific language-learning problems as well as the child's level of modifiability to mediated language experiences.

Child B. Child B was a female aged 4:6 who spoke mainly Spanish at home and school. She was exposed to English at school and in her community and was able to speak and understand some words in English. Her teachers had no concerns about her language skills. Classroom interaction showed that she used words, phrases, and sentences to initiate conversations with, and respond to, peers. She followed directions and gave commands in group play in the "dress-up" area and was able to maintain conversations for approximately three exchanges with peers.

Child B's EOWPVT-R standard score of 71 was average compared to those of her same-age Head Start peers, but 1.93 SD below the national norm; her CSSB standard score of 100 was average compared to the national norms ($M = 100$, $SD = 16$) and the local norms ($M = 91.90$, $SD = 14.47$). On the PLS-A, she achieved 7 items correct out of 10 possible items, which was within 1 SD for the peer comparison group.

Overall, it appeared that Child B did not have language deficits even though her EOWPVT-R score was approximately 2 SD below the mean compared to the national norms. If the measures used were applied to identify language deficits, the classroom observations, teacher reports, and comparisons to the available local norms helped rule out the presence of potential language problems in this child. For Child B, performance compared to the national norms revealed a language difference, not a language deficit.

Some limitations of existing measures. For many children, classroom observations and teacher reports combined with comparisons to local norms may enable clinicians to rule out a language disorder. Yet, these approaches have limitations that may not allow for the differentiation of language disorders in some cases. First, classroom observations and teacher reports may not correspond to children's scores on tests because they represent general estimates of the child's performance. Second, sampling variables may impact the validity and reliability of these observations as well. Observations may be useful to rule out language impairment only in the context of other sources of converging evidence. Similarly, teacher reports of concern may be helpful to identify children who may need further assessment, but they may

not be sufficient for establishing the nature of the child's difficulties or how to address the child's needs. Finally, even if language tests can be developed for different CLD groups with appropriate local norms as well as acceptable diagnostic sensitivity and specificity, this type of static assessment does not address individual differences in previous educational opportunity or experience. In addition, this type of static assessment does not provide information about the child's learning ability or ways to facilitate the child's language performance for those children with specific needs.

Knowing that Child A and Child B scored similarly to other same-age CLD children on a targeted test is not useful to address the child's language differences that affected test performance in the first place, and that may result in low performance compared to majority-culture children on other types of language-based assessments in the future. The use of local norms to make educational decisions also raises serious ethical questions because they tend to generate significantly lower expectations for CLD children as compared to majority-culture children.

As discussed earlier, the purpose of the test-teach-retest model is to *modify* the child's performance by teaching the principles that underlie the task. Within this DA model, the role of the examiner is to induce change that can generalize to other tasks. Thus, the critical component of this model is to address those same language differences that affected initial test performance and, ultimately, the goal is to equalize differences in performance across diverse groups. The focus on equalizing differences in past experience versus using assessments to simply "qualify" a child for services has far-reaching implications. Hart and Risley (1995) argued that, by addressing these differences in cultural experience, we will be able to facilitate children's access to advanced educational and occupational opportunities as adults. Although the interventions used in DA may not be sufficient to provide children with needed MLEs, they may allow clinicians to assess the child's ability directly in order to profit from such experiences.

Table 1 summarizes standardized and PLS-A raw scores for the two children pre- and post-mediation. The PLS-A

scores ranged from 4 to 7, with little differential performance between the children as compared to mean scores for the local group ($M = 4.00$, $SD = 2.84$). This was not surprising given the limited number of items tested and the fact that they assessed school-based concepts, which the children were beginning to learn. The test results provided no information concerning these children's modifiability for these tasks.

For Child A, it was difficult to know whether discrepancies in test performance related to specific language-learning difficulties or only to differences in experiences with the language tasks. For Child B, it was important to know the extent to which her performance, compared to national norms, was related to differences in previous experiences, and whether those differences could be addressed with examiner mediation. As discussed earlier, answers to these questions are critical because CLD children may be at risk if their language differences are not met. Concluding that it was "typical" for some of these children to score 2 SD below the national norm was not sufficient. Thus, the goal of the DA was not limited to examining children's ability to change performance on the targeted measures. Rather, the goal included determining what was needed to induce change that would maximize these children's potential for success with other language tasks in the future. The DA model illustrated here addresses these two issues.

Training Phase and Measures of Language Change

The MLE consisted of teaching children strategies for using single-word labels during two 30-minute sessions, 2 weeks apart. Specifically, the examiner used the MLE strategies of intentionality, mediation of meaning, transcendence, and competence to teach the children about using "special names." Intentionality was used to convey to the children that the focus of the session was to learn about special names. Mediation of meaning was used to explain to children why it was important to use single-word labeling. Transcendence was used to help children understand what

Table 1. Dynamic assessment of single-word labeling for Child A and Child B pre- and posttesting.

Child	Age	Gender	Teacher concern	EOWPVT-R		CSSB		PLS-A	
				M 100	SD 15	M 100	SD 16	RS Pre-	RS Post-
				SS Pre-	SS Post-	SS Pre-	SS Post-	RS Pre-	RS Post-
A	4:0	F	Y	67	67	78	86	4	3
B	4:6	F	N	71	86	100	116	7	7
Group means ^a (Standard deviation)				72.76 (8.08)	83.56 (10.02)	94.05 (12.71)	99.24 (13.26)	4.41 (2.74)	7.10 (2.02)

Note. EOWPVT-R = Expressive One-Word Picture Vocabulary Test-Revised; CSSB = Comprehension subtest of the Stanford-Binet Intelligence Scale (4th ed.); PLS-A = Preschool Language Scale (adapted version); M = mean; SD = standard deviation; SS = standard score; RS = raw score.

^a Group means correspond to local norms derived from typically developing Head Start peers before and after mediation (Peña et al., 2001).

would happen if specific single-word labels were not used, and to discuss situations at home and in the classroom that might require the use of labels. Table 2 provides examples of how these strategies were used. Finally, for mediating competence, children were encouraged to think about the strategies they would use to label and when the application of these strategies might be important.

The DA measures included modifiability ratings, language gain scores, and an analysis of the children's responses to the tests post-MLE. Two rating scales were used during the MLE to rate child performance within and across the two sessions (see Appendix). The Learning Strategies Checklist (Peña, 1993) was used to rate attention, comparative behavior, planning, self-regulation, transfer, and motivation of the children within each session (Learning Strategies Checklist: LSC-1 and LSC-2). There were 16 ratings, ranging from 0 to 2 (none of the time, some of the time, most of the time) for a possible total summary score between 0–32 for each of the two sessions. Second, the Modifiability Scale (Peña, 1993, based on Lidz, 1991) was used to summarize the child's overall modifiability to the MLE. Modifiability was rated using Likert scales based on the following components: examiner effort (0–3 points), child responsivity (0–3 points), and transfer (0–2 points). To determine children's language gains post-MLE, children were retested using the EOWPVT-R, CSSB, and PLS-A within 6–8 weeks of the pretest.

Table 2. Mediated learning experience strategies, definitions, and examples.

Strategy	Definition ^a	Example ^b
Intentionality	The mediator focuses both visual and verbal attention of the learner and communicates expectation of success.	"Today, we're going to think about special names."
Meaning	The mediator facilitates the learner's awareness of learning targets by attributing meaning and value to specific behavior.	"Special names help us tell things apart."
Transcendence	The mediator makes explicit the expectation that the newly learned skills will be used across many content areas. Examples are used to bridge learning tasks to the child's experiences.	"What would happen if I called you [wrong name]? Would you know who I was talking to?"
Competence	Reinforcement is used to encourage the use of new strategies and to provide motivational support for attempting new and unfamiliar tasks or skills.	Good! At first you didn't use special names, but we worked on that and now you know that special names are important.

^a Adapted from Feuerstein (1980); ^b See Peña et al. (2001) for additional examples.

Evaluating Individual Profiles

Table 3 shows the LSC-1, LSC-2, and Modifiability Scale scores obtained by the two children. Table 1 presents the language scores pre- and post-MLE for the two children. The pre-/posttest comparisons revealed important information concerning the children's ability to generalize the new strategies that were taught during the MLE to the test situation. The analysis of their language-learning behaviors during the MLE also showed different patterns of performance.

Child A. Child A demonstrated high motivation and attention during the MLE sessions. She inconsistently used strategies of planning and self-regulation during the sessions. She practiced the object labels presented during the sessions and asked the examiner for clarification when she was not sure. This demonstrates awareness of the task goals and of her performance on the task. However, her profile demonstrated very modest changes over time across tests. She made no change in single-word labeling, which was the focus of the MLE sessions (pretest = 67, posttest = 67). Her posttest score placed her 1.65 *SD* below the mean of her Head Start peers who had participated in similar mediation sessions. Although she made changes on the CSSB (pretest = 78, posttest = 86), her posttest score still placed her 1 *SD* below the mean score obtained by her peers after mediation. Her posttest changes on the PLS-A were insignificant (pretest = 4, posttest = 3) and, in comparison to her peers who received similar mediation experiences, her score of 3 was 2.51 *SD* below the group mean.

An analysis of Child A's responses to specific items showed that she provided more elaborate responses on the CSSB posttest as compared to the pretest (3–4 words vs. 6–8 words), and that the responses were more related to the question, although many of these answers continued to be incorrect. For example, in response to "What do people do when they're thirsty?" Child A responded "no se" (I don't know) on the pretest, and "Se meten en el agua y juegan" (They get in the water and play). She was moderately responsive to the MLE within the session, with an apparent weakness in her ability to transfer learned information across tasks. This suggests a need for intervention to facilitate learning to generalize strategies to similar and new tasks.

These results demonstrate continued needs in the area of language, and Child A likely would be diagnosed with a language-learning problem. Suggestions for intervention approaches include helping her to develop a plan for completing tasks and helping her to understand her correct and incorrect responses. Intervention within language therapy and the classroom setting should make the most of her attention to task and her high persistence and enthusiasm.

Child B. Child B demonstrated high responsiveness during the MLE sessions, obtaining high scores in the areas of examiner effort, responsivity, and transfer. During the MLE, she demonstrated strengths in the areas of attention, planning, self-regulation, transfer, and motivation. Child B initiated and maintained focus on task and made comments about features of the task related to the goal of labeling.

Table 3. Dynamic assessment of single-word labeling for Child A and Child B mediated learning experience results.

Child	LSC-1	LCS-2	Examiner effort	Responsivity	Transfer
	RS (0-26)		RS (0-3)	RS (0-3)	RS (0-2)
A	17	23	2	2	1
B	20	25	3	3	2

Note. LSC = Learning Strategies Checklist; 1 = first mediated learning experience session; 2 = second mediated learning experience session; RS = raw score.

She talked about the goal of “special names.” She waited for instructions from the examiner and asked for help when needed. As Child *B* learned about special names, she evidenced self-correction behavior. She obtained the highest possible modifiability score of 8, indicating that minimal examiner effort was required for her to make changes. She was highly responsive to intervention and she transferred learned skills across tasks and between sessions.

Child *B*’s pretest profile demonstrated relative weakness in the area of single-word naming (EOWPVT-R standard score = 71), with strengths in descriptive skills (CSSB standard score = 100) and academic skills (PLS-A raw score = 7). Her performance after mediation showed improvement on both the EOWPVT-R ($SS = 86$) and the CSSB ($SS = 116$), with no change on the PLS-A. The EOWPVT-R and PLS-A scores placed her within average range in comparison to the local norms at posttest; her CSSB posttest score was 1.26 *SD* above the posttest group mean. Posttest performance was consistent with her modifiability profile, as she improved performance on the EOWPVT-R and CSSB, indicating near and far transfer. She showed no change on the PLS-A; however, she had a high score (7/10) pre-MLE. As suspected, given her overall profile and her ability to make changes with adult mediation, Child *B* did not require specific intervention. In this case, DA was used to provide her with more experiences with the task of single-word labeling, which enabled her to boost her score significantly and to demonstrate her true vocabulary abilities on this task.

These examples illustrate a test-teach-retest approach that can help differentiate language differences from disorders. The examples also provide useful information regarding the nature of children’s limited language performance and the type of interventions that may be needed to facilitate change. For example, Child *B* initially demonstrated limited performance on the EOWPVT-R, but with DA, she was able to demonstrate high levels of performance, thereby corroborating the initial presence of language differences. For Child *A*, the MLE demonstrated inconsistent use of language strategies, and posttest results were consistent with this observation in that no improvements were observed on the EOWPVT-R. Based on those results, an intervention plan that capitalized on demonstrated motivation and attention skills while working on weaker planning and self-regulation strategies would be developed.

CONCLUSIONS

The two case studies and other examples discussed earlier in the article demonstrate how a test-teach-retest approach could be used effectively to differentiate language differences from disorders in CLD children across various language areas (e.g., vocabulary, narrative, synonyms/antonyms). Unlike “teach the test” strategies, the intervention phase of the DA does not use the actual test items or materials, but instead provides sufficient flexibility to address children’s individual differences in performance by varying the activities in order to facilitate learning.

The measures of change, such as gain scores, ratings of modifiability, and qualitative changes, may be extremely useful for differentiating language differences from disorders and for determining appropriate educational decisions. Quantitative data, such as test scores at pretest and posttest, can provide information concerning children’s ability to make changes as a result of a short-term intervention, such as that demonstrated in the case studies. The expectation is that children who score low at pretest but who have normal language-learning ability would show significant change. Children with language-learning problems who also may score low at pretest, and who may not be differentiated from their typical peers on a standardized test such as those used in the case studies, may make limited or no changes after the MLE.

Pre-/posttest comparisons have limitations, however, particularly for individual examinees, because such comparisons are less reliable than the reliabilities of pre- or posttest scores alone (Allen & Yen, 1979). In the case studies discussed earlier, there were differences between Child *A*, who made significant gains after mediation but was still scored low on the posttest, and Child *B*, whose gains were high enough to place her at or above the group mean on the posttest score. Gain scores were not as revealing of the children’s skills as were their posttest scores.

Furthermore, gain scores are not equal across the same test. For example, the difference between a score of 10 on the pretest and 15 on the posttest may not be equivalent to a 5-point difference between a pretest score of 40 and a posttest score of 45 because of differences in item difficulty (Embretson, 1987). Although the use of standard scores may provide a more stable measure of change, scores in the central portion of the normal curve are more

stable than those at the tails (Lord, 1980). For children who are performing near their ability level, it may be difficult to push their scores beyond the pretest score. In contrast, for a child who is performing somewhat below level but who has normal ability, a significant change would be expected. Thus, clinicians should use posttest scores (see Embretson, 1991, for a discussion) rather than gain scores in conjunction with their observations of modifiability to report the results of their DAs.

Clinicians also should describe changes in the quality of children's responses as additional measures of change. Qualitative changes provide specific information about areas to target in intervention for a given child. For example, Child A showed changes in response type (e.g., more elaborated responses on the posttest in comparison to "I don't know" responses on the pretest), even though her posttest responses were still incorrect. These measures also are useful for establishing the child's level of readiness for learning a specific language behavior. Furthermore, information about how children respond during the MLE session can be useful for determining underlying strategies that impact language learning. Information derived from the modifiability scales discussed in this article can be useful for incorporating other learner variables (e.g., attention, motivation, etc.) in making clinical recommendations. Children who have not had specific educational experiences may develop feelings of incompetence and a lack of motivation, which may not allow them to benefit from traditional instruction. Thus, the scales can help clinicians identify differences that relate to limited educational opportunity (rather than a disorder). The scales also are helpful to determine what it takes to effect change, such as the amount of effort/intensity, time, or type of cue for modifying children's performance.

When an appropriate MLE is provided, children who are different, but typical, language learners are capable of demonstrating significant changes. On the other hand, children with language impairments may benefit from the MLE sessions but would demonstrate little or no quantitative change. For these children, the assessment can provide useful information about the child's future responsiveness to intervention for the specific language areas targeted. Although there are still issues that need to be addressed, the DA approach discussed in this article should help clinicians reduce test bias and make better clinical decisions for all children.

ACKNOWLEDGMENT

This work was supported in part by a grant from the National Institutes of Health (National Institute of Deafness and Other Communication Disorders) #5 K23 DC00141 to Dr. Elizabeth Peña.

REFERENCES

- Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth.
- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist*, 36(10), 1086–1093.
- 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), 81–92.
- Budoff, M. (1974). *Learning potential and educability among the educable mentally retarded*. (Final Report Project No. 312312). Cambridge, MA: Research Institute for Educational Problems, Cambridge Mental Health Association.
- Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 173–195). New York: Guilford.
- Campione, J., & Brown, A. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). New York: Guilford.
- Carlson, J., & Wiedl, K. (1980). Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical formulations. *Zeitschrift für differentielle und diagnostische psychologie*, 1(4), 303–318.
- Carlson, J., & Wiedl, K. H. (1978). Use of testing-the-limits procedures in the assessment of intellectual capabilities in children with learning difficulties. *American Journal of Mental Deficiency*, 82, 559–564.
- Carlson, J., & Wiedl, K. H. (1992). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 167–186). New York: Springer-Verlag.
- Embretson, S. E. (1987). Toward development of a psychometric approach. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 141–170). New York: Guilford.
- Embretson, S. E. (1991). Implications of a multidimensional latent trait model for measuring change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 184–197). Washington, DC: American Psychological Association Books.
- Evans, L., Maschmeyer, J., & McFarlane, L. (1996, November). *Relationship between awareness and severity of phonological disorders*. Paper presented at the Annual Meeting of the American Speech-Language-Hearing Association, Seattle, WA.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore, MD: University Park Press.
- Feuerstein, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore, MD: University Park Press.
- Fey, M. (1986). *Language intervention with young children*. Austin, TX: PRO-ED.
- Gardner, M. F. (1979). *Expressive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- Gardner, M. F. (1990). *Expressive One-Word Picture Vocabulary Test-Revised*. Novato, CA: Academic Therapy Publications.
- Ginsburg, H. P. (1986). Academic diagnosis: Contributions from developmental psychology. In J. Valsiner (Ed.), *Individual subject and scientific psychology* (pp. 235–260). New York: Plenum.
- Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. New York: Cambridge University Press.

- Gray, S., Plante, E., Vance, R., & Henrichsen, M.** (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30*, 196–206.
- Gutiérrez-Clellen, V. F., Brown, S., Robinson-Zañartu, C., & Conboy, B.** (1998). Modifiability. *Journal of Children's Communication Development, 19*(2), 31–43.
- Gutiérrez-Clellen, V., Peña, E., & Quinn, R.** (1995). Accommodating cultural differences in narrative style: A multicultural perspective. *Topics in Language Disorders, 15*(4), 54–67.
- Gutiérrez-Clellen, V. F., & Quinn, R.** (1993). Assessing narratives in diverse cultural/linguistic populations: Clinical implications. *Language, Speech, and Hearing Services in Schools, 24*(1), 2–9.
- Hart, B., & Risley, T. R.** (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brookes.
- Jedrysek, E., Klapper, A., Pope, L., & Wortis, J.** (1972). *Psychoeducational evaluation of the preschool child*. New York: Grune and Stratton.
- Lidz, C.** (1996). Dynamic assessment and the legacy of L. S. Vygotsky. *School Psychology International, 16*(2), 143–154.
- Lidz, C. S.** (1987). Historical perspectives. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 3–34). New York: Guilford.
- Lidz, C. S.** (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford.
- Lidz, C. S., & Thomas, C.** (1987). The preschool learning assessment device: Extension of a static approach. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 288–326). New York: Guilford.
- Long, S. H., & Olswang, L. B.** (1996). Readiness and patterns of growth in children with SELI. *American Journal of Speech-Language Pathology, 5*(1), 79–85.
- Lord, F.** (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Miller, L., Gillam, R., & Peña, E.** (2001). *Dynamic assessment and intervention: Improving children's narrative abilities*. Austin, TX: PRO-ED.
- Norris, J., & Hoffman, P.** (1990). Language intervention within naturalistic environments. *Language, Speech, and Hearing Services in Schools, 21*, 72–80.
- Olswang, L., & Bain, B.** (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech and Hearing Research, 39*, 414–423.
- Olswang, L., Bain, B., & Johnson, G.** (1992). Using dynamic assessment with children with language disorders. In S. F. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 187–216). Baltimore, MD: Paul H. Brookes.
- 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.** (1996). Dynamic assessment: The model and language applications. In K. Cole, P. Dale, & D. Thal (Eds.), *Assessment of communication and language* (pp. 281–307). Baltimore, MD: Paul H. Brookes.
- Peña, E.** (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly, 21*(2), 87–97.
- Peña, E.** (2001). Assessment of semantic knowledge: Use of feedback and clinical interviewing. *Seminars in Speech and Language, 22*(1), 51–63.
- Peña, E.** (n.d.). [Performance of African American and Puerto Rican preschoolers on the EOWPVT in feedback vs. no-feedback conditions.] Unpublished raw data.
- Peña, E., Iglesias, A., & Lidz, C.** (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*, 138–154.
- Peña, E., & Quinn, R.** (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools, 28*, 323–332.
- Peña, E., Quinn, R., & Iglesias, A.** (1992). The application of dynamic methods to language assessment: A non-biased procedure. *Journal of Special Education, 26*, 269–280.
- Reyes, M.** (1992). Challenging venerable assumptions: Literacy instruction for linguistically different students. *Harvard Educational Review, 62*, 427–446.
- Spector, J.** (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84*, 353–363.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M.** (1986). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago, IL: Riverside.
- Tzuriel, D., & Klein, P.** (1985). Analogical thinking modifiability in disadvantaged, regular, special education, and mentally retarded children. *Journal of Abnormal Child Psychology, 13*, 539–552.
- Tzuriel, D., & Klein, P.** (1987). Assessing the young child: Children's analogical thinking modifiability. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 268–287). New York: Guilford.
- Vygotsky, L. S.** (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S.** (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Zimmerman, I., Steiner, V., & Pond, R.** (1978). *Preschool Language Scale*. Columbus, OH: Charles E. Merrill.

Received December 12, 2000

Accepted May 24, 2001

DOI:10.1044/0161-1461 (2001/019)

Contact author: Vera F. Gutiérrez-Clellen, Department of Communicative Disorders, San Diego State University, San Diego, CA 92182-1518. Email: vcllellen@mail.sdsu.edu

APPENDIX. RATING SCALES

Learning Strategies Checklist

	None of the time	Some of the time	Most of the time
Attention/Discrimination			
• initiates focus with minimum cues	0	1	2
• maintains focus with minimum cues	0	1	2
• responds to relevant cues, ignores irrelevant cues	0	1	2
Comparative Behavior			
• comments on features of task	0	1	2
• uses comparative behavior to select item	0	1	2
• talks about same/different	0	1	2
Planning			
• talks about overall goal	0	1	2
• talks about plan	0	1	2
Self-Regulation/Awareness			
• waits for instructions	0	1	2
• seeks help when difficult	0	1	2
• corrects self	0	1	2
• rewards self	0	1	2
Transfer			
• applies strategies within tasks	0	1	2
• applies strategies between tasks	0	1	2
Motivation			
• persists even when frustrated	0	1	2
• shows enthusiasm	0	1	2

© Copyright 1993 by E. Peña. Reprinted with permission.

Modifiability Scale

	Extreme	High-Moderate	Moderate	Slight
Examiner Effort	3	2	1	0
Child Responsivity	3	2	1	0
	Low	Medium	High	
Transfer	0	1	2	

© Copyright 1993 by E. Peña. Reprinted with permission.