Research

Reducing Test Bias Through Dynamic Assessment of Children's Word Learning Ability

Elizabeth Peña
University of Texas at Austin

Aquiles Iglesias
Temple University, Philadelphia, PA

Carol S. Lidz
Touro College, New York, NY

This study examined the performance of preschool children from culturally and linguistically diverse backgrounds, both typically developing and with low language ability, on a word-learning task. A pretest-teach-posttest method was used to compare a mediation group to a no-mediation group. Children in the mediation group were taught naming strategies using mediated learning experience (MLE). Results indicated that typically developing and low language ability children were differentiated on the basis of pretest-posttest change and that dynamic measures (e.g., posttest scores of single-word labeling and modifiability ratings from the mediation sessions) predicted the ability groups better than static measures (e.g., pretest scores of single-word labeling, description, and academic concepts). These results suggest that dynamic assessment approaches may effectively differentiate language difference from language disorder.

Key Words: dynamic assessment, mediated learning experience, nonbiased assessment, culturally and linguistically diverse, preschool

Despite the implementation of "nonbiased" assessment practices in many educational settings, children from culturally and linguistically diverse (CLD) backgrounds continue to be misdiagnosed by professionals in special education, including speech-language pathologists (Gersten & Woodward, 1994; Yates, 1988). In the area of language, children may be misdiagnosed as a result of different home and school language experiences (Blount, 1982; Heath, 1983, 1986; Rogoff, 1991; Schieffelin & Ochs, 1986). These differences have the potential of influencing standardized test performance. Research demonstrates that, whereas children from nonmainstream groups are competent language users in their speech communities, they may not be viewed as competent in other contexts (e.g., school or assessment contexts) (Heath, 1983; Iglesias, 1985). Specifically, the previous experiences these children bring to school may not match teacher or test demands, resulting in low performance in classrooms and on standardized achievement tests, formal language tests, and psychological tests (Bain & Dollaghan, 1991; Bereiter, 1985; Cummins, 1984, 1986; Duran, 1989; Heller, Holtzman, & Messick, 1982; Wright & Santa Cruz, 1983).

An area where variation is often seen across CLD children is vocabulary acquisition. This variation in lexical acquisition is closely related to the way parents interact with their children (Della Corte, Benedict, & Klein, 1983; Tomasello & Farrar, 1986; Tomasello & Todd, 1983). Peña and Quinn (1997) suggested that these different types of interaction styles might influence the knowledge and skills necessary for adequate test performance. For example, in her analysis of Puerto Rican mother-child interaction, Quinn (1992) found that mothers used very few nouns in their interactions with their 12-month-old infants. Indeed, error analysis of Puerto Rican preschool children's performance on a labeling task revealed use of descriptive responses, substitutions, or related responses (Gutierrez-Clellen & Iglesias, 1987). Similarly, Heath (1983) noted that, among African-Americans in the rural South, adults rarely elicited single-word nouns from preschool children, but they frequently sought comparisons, explanations, nonverbal responses, and information responses (in this order). Anderson-Yockel and Haynes (1994) reported that working-class African American mothers rarely asked questions, and their children produced

more spontaneous verbalizations. Thus, a test task that requires naming single-word labels in response to adult questions may instead elicit longer responses such as explanations or descriptions.

Given the possible diversity of language performance of children from various cultural and linguistic backgrounds, low performance on a single administration of a test that elicits single-word responses to adults' questions may not be indicative of language disorder. Alternatively, low performance of a child from a CLD background may not necessarily reflect language difference. Language assessment should examine initial test performance as well as language learning ability in order to rule out possible test bias. Dynamic assessment provides an alternative approach to traditional procedures by focusing on learning potential, possibly reducing test bias. This study explores the use of dynamic assessment for differentiating language difference from language disorder in children from diverse cultural and linguistic backgrounds.

Dynamic Assessment

Major characteristics of the dynamic assessment model include a test-teach-retest format and a focus on the learning process. The theoretical grounding for dynamic assessment is Vygotsky's notion of the "zone of proximal development" (ZPD; Vygotsky, 1986). Vygotsky criticized Western approaches to psychological assessment for relying solely on estimation of the child's independent performance (zone of actual development) without attention to the child's ability to profit from instructional interaction with more experienced collaborators. He considered reliance on independent functioning both incomplete and inaccurate, as children with the same estimated levels of unassisted performance can be quite different when involved in a learning experience. Further, Vygotsky suggested that development of higher mental functioning required social interaction within zones of proximal (next) development. Therefore, if assessment is to have validity for learning situations that aim to promote higher levels of mental functioning, it should include such interactions in order to sample the child's functioning within the ZPD (Minick, 1987). Feuerstein (1979), Lidz (1987, 1991, 1995) and others have applied Vygotsky's ideas about promoting higher functioning within the ZPD in descriptions of the mediation interaction that occurs during the teaching phase of dynamic assessment.

During this mediation phase of dynamic assessment, evaluators engage children in mediated learning experiences (MLE) that are designed to teach children about the principles of task solution or problem-solving strategies that underlie successful test-taking performance. Observed changes in the child's learning strategies and development of competence in response to MLE are used to describe learning ability, or ability to profit from instruction (Feuerstein, 1979; Haywood & Wingenfeld, 1992; Lidz, 1987, 1991). Thus, dynamic assessment seeks to assess learning within the child's ZPD.

According to Feuerstein, four mediation components are critical for an MLE: intentionality, transcendence, meaning, and competence. *Intentionality* refers to the mediator's intent to teach, and it is made explicit by a statement about the goal for the lesson. The mediator engages the child in the interaction for the purpose of teaching and helps the child know the purpose of the interaction. Transcendence is believed to help children learn to think hypothetically (Lidz, 1991). *Transcendence* mediations include linking the immediate task to events in the child's experience and asking "what if" questions during joint problem-solving activities. Highlighting the importance and features of the lesson transmits *meaning* to the child. Through focus on meaning, the child's attention is directed to the critical features of the task, and the mediator enhances the child's awareness of what to notice and why the task is relevant. Finally, helping the child develop, plan, and carry out strategies for approaching a task fosters *competence*. The ultimate goal in MLE is to help the child become a self-regulated, active learner and symbolic thinker (see Lidz, 1991 for a more complete review of MLE).

Research using MLE as a diagnostic indicator has focused on the differential responses on cognitive and reading measures between mediated and nonmediated groups. These studies have used MLE in a one-time, pretest-teach-posttest paradigm. Studies with typically developing preschool children (Lidz & Thomas, 1987), children with learning disabilities (Missiuna & Samuels, 1989), and preschoolers with developmental delay (Reinharth, 1989) have demonstrated higher performance for children receiving MLE compared to those in a no-mediation group. The results from these studies suggest that MLE helps children to perform in their zone of proximal development. Because MLE focuses on giving children the foundation for competence and prerequisite experience with targeted tasks, researchers have suggested that dynamic assessment with MLE can be used to reduce test bias (Gutierrez-Clellen, Peña, & Quinn, 1995; Gutierrez-Clellen & Quinn, 1993).

Experimental studies have demonstrated that children with language impairment have more difficulty learning targeted language features (e.g., Ellis Weismer & Hesketh, 1993; Kiernan & Gray, 1998). If application of dynamic assessment can lessen the effects of lack of previous experience on test performance, then low-scoring children with and without language impairment should show differential pretest-posttest performance. Budoff (1987), for example, applied his dynamic assessment procedures to more accurately classify children for special education. He proposed that children who demonstrated limited change in dynamic assessment were those who needed more intensive intervention through special education placement.

Peña, Quinn, and Iglesias (1992) used a similar pretest-posttest approach to demonstrate the application of dynamic assessment as a nonbiased method for testing vocabulary skills in Puerto Rican and African-American preschool children. Their pretest and posttest measure was the Expressive One-Word Picture Vocabulary Test (EOWPVT; Gardner, 1979). Although there were no significant pretest score differences between children with and without possible language disorder, pretest-posttest patterns differentiated children by language ability. Furthermore, observations of modifiability using Likert

scales—ratings of examiner effort, child responsiveness to teaching, and the child's ability to transfer information to a new context—differentiated typically achieving and low language ability children. However, this study lacked a no-mediation group, so it is uncertain whether pretest-posttest differences were due to dynamic assessment or whether increased exposure to the test or the school curriculum would yield similar patterns for the two ability groups. Following up on the EOWPVT pretest results from the 1992 study, Peña and Quinn (1997) compared the performance of Puerto Rican and African American children of different language abilities on test tasks that entailed labeling and description. Their findings indicated that children with typical and low ability performed differently on the description task, but that there were no ability differences on the labeling task.

The above discussion suggests a need to further explore the findings of Peña and colleagues in order to determine whether dynamic assessment, using a pretest-MLE-posttest approach, could improve assessment accuracy in comparison to a no-mediation group. Specifically, we examined whether a dynamic assessment procedure could differentiate children with typical versus low language ability in comparison to static (pretest measures). A no-mediation group was added to enable us to account for possible maturation as well as school exposure effects. We also included a number of test tasks to help evaluate whether dynamic assessment (i.e., MLE) effects transfer to performance in related areas (Haywood & Wingenfeld, 1992; Lidz, 1996). We investigated pretest-posttest performance of African-American and Latino-American children with typical and low language development. The present study asks the following questions: (a) Do children who are typically developing and children with low language ability respond differently to MLE? and (b) How well do dynamic assessment procedures, compared to static assessment procedures, classify preschool children from diverse CLD backgrounds by language ability?

Method

Participants

Seventy-nine children, ages 3;9 to 4;9 (mean age = 4;2), enrolled in a bilingual Head Start program in an urban Northeastern city, participated in the study. All were reported to be in good health and had passed a hearing screening conducted by the Head Start program. Children were recruited from five full-day classes at two different centers located within walking distance of each other. There were 20 children in each class, and all age-eligible children were invited to participate in the study. As a result of programmatic constraints, children were not randomly assigned to mediation groups. Thus, children from Center A served as the experimental (MLE) group and children from Center B served as the no-mediation group. Both centers had similar distributions in terms of ethnicity and home language as reported by the parents, although there were different gender distributions. Participants included 55 children from Center A (Class 1: 20 children, Class 2: 16 children, and Class 3: 19 children) and 24 children from Center B (Class 4: 11 children and Class 5: 13 children). Table 1 displays the demographic characteristics for the two centers. Each class in the two centers had a teacher and a teacher assistant, at least one of whom was bilingual (Spanish/English). The classrooms followed the same curriculum at the same pace, as established by the education coordinator who served both centers. A single bilingual model was not implemented across classes, resulting in varied uses of Spanish and English. Children were nevertheless exposed to both languages and the same materials and curriculum, making it unlikely that these factors had a significant impact on the outcomes of the present study.

One goal of the study was to report posttest performance in response to MLE regardless of language proficiency. Therefore, testing and MLE were conducted in the language(s) of the child. For children who responded in both languages, conceptual scoring (based on Pearson,

TABLE 1. Participant demographics.

	Center A MLE Group		Center B No-Mediation Group	
	Demographic	Percent of Group	Demographic	Percent of Group
Mean age (and *SD*) in months	50.4 (2.96)		50.0 (3.23)	
Children	55		22	
Latino children	43	78.20	17	77.30
African American children	12	21.80	5	22.70
Speakers				
Monolingual English speakers	29	52.70	10	45.50
Bilingual Spanish/English speakers	20	36.40	11	50.00
Monolingual Spanish speakers	6	10.90	1	4.50
Gender				
Girls	19	34.50	14	63.60
Boys	36	65.50	8	36.40
Language ability grouping				
Low language ability	10	18.20	5	22.70
Passed 0 criteria	1	1.80	0	0.00
Passed 1 criteria	3	5.50	1	4.50
Passed 2 criteria	6	10.90	4	18.20
Typical development	45	81.80	17	77.30
Passed 3 criteria	4	7.30	2	9.10
Passed 4 criteria	11	20.00	1	4.50
Passed 5 criteria	30	54.50	14	63.60

Note. Criteria for language ability included sum total number of measures passed (of a total of 5) to yield the "Language Ability Composite." These measures included the Brigance Inventory of Early Development, teacher observation, parent report, classroom observation of peer initiations, and classroom observation of peer responses. Children who passed 3, 4, or 5 criteria were identified as TD; children passing 0, 1, or 2 criteria were identified with LLA.

Fernandez, & Oller, 1993)—where both Spanish and English items are counted, but counted only once per item—was used to estimate the raw scores. For example, a child responding *dog*, both *bird* and *pájaro* (bird), and *pato* (duck) to pictures of a dog, a bird, and a duck would receive a score of 3. This type of scoring helped to estimate language concepts across languages without under- or overestimating test scores. Thus, the heterogeneity of participants' language backgrounds and language use did not interfere with the purpose of the study. Other possible differences (ethnicity, gender, and home language) were explored statistically and are reported below.

External Validity Measures

There are no established standardized procedures for appropriate language assessment of bilingual children and children from culturally and linguistically diverse backgrounds. There are, however, guidelines in the literature, which recommend the use of naturalistic, ecologically valid procedures. Thus, children were identified with low language ability or typical development based on a combination of measures including: classroom observation (Damico, 1990; Fey, 1986; Hamayan & Damico, 1990; Kovarsky & Maxwell, 1992; Westby & Erickson, 1992), parent report (Bates, Bretherton, & Snyder, 1991; Dale, Bates, Reznick, & Morisset, 1989; Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993; Vincent, 1992), and teacher report (Parnell, 1994). The classroom observation, parent reports, and teacher reports used to identify child language ability served as an ecologically validated criterion for evaluation of the dynamic assessment procedure. Children with typical development met at least three of the following criteria, whereas children with low language ability failed at least three of the same criteria:

1. Parent reports of children's home language use and comprehension indicated that the parents judged their child's language performance to be appropriate.
2. Teacher-administered Brigance Inventory of Early Development (Brigance, 1978) indicated that children performed appropriately for their age and sociocultural group.
3. Classroom teacher judgments of classroom language use and language comprehension indicated that they had no concerns about language development.
4. During a 10-minute observation of play with peers, children initiated interaction using complete sentences.
5. During a 10-minute observation of peer interaction during play, children responded to peers using complete sentences.

Using these criteria, 11 children from Center A and 6 children from Center B were identified as having low language ability (LLA) (see Table 1). The 20–25% incidence of LLA in the participant pool is greater than the incidence in the general population. However, there are factors that may explain the reported rate. First, Head Start has a mandate to set aside 10% of their slots for children with disabilities. Children with suspected language impairment fill most of these slots. Enrollment procedures for the additional 90% of the slots do not exclude children with disabilities, so it is likely that some of these children will also demonstrate unidentified language impairment. Second, our purpose was to identify children with low language ability relative to the total participant pool, but this does not mean that LLA is equivalent to language impairment. It is likely that within the LLA group there were children with language impairment as well as those with language ability in the low-normal range.

Procedures used to collect these data ensured that the graduate students working directly with the children would be blind to language status (LLA vs. TD) in the following ways. First, parent, teacher, and observation data were not scored until the completion of the study. Furthermore, data were collected on a rotation basis so that clinicians gathering baseline data for the children from one classroom were assigned to another classroom for the experimental portion of the study.

Collection of Parent Reports. During the summer before school began and the first 2 weeks of enrollment, program staff collected parent reports regarding possible concerns about language development as part of their routine developmental and health history intake. Questions regarding language development asked whether another agency had identified the child as having language impairment and whether the parents believed their child's language development was progressing at a normal rate. Based on these reports, a binary (pass = no concerns, fail = concerns) score was recorded for each participant (coded as 0 and 1, respectively).

Collection of Teacher Reports. Once the children had been attending school for 4 weeks, teachers were individually interviewed regarding their observations of the children's classroom language performance. Going through the list of enrolled children in the classroom, teachers were asked whether they judged children's expressive language, receptive language, or articulation to be age-appropriate. When teachers indicated that they were concerned about a particular child, they were asked to give examples for illustration. If the teachers were concerned about expressive or receptive language, a score of "fail" (later coded as 0) was assigned. If the teachers indicated no concerns, a score of "pass" (coded as 1) was assigned. Articulation errors, in the context of no language concerns, were scored as "pass."

Classroom Observation. Observation of peer interaction based on Fey's (1986) model of responsiveness and assertiveness was used to evaluate communicative competence. Bilingual graduate students observed each child interacting with peers during a 10-minute free play session. Clinicians identified the mode (sentence, phrase, single-word, gesture, or vocal) and appropriateness (coded +/–) of the child's initiations with peers and the child's responses to peers. Separate summary scores were calculated for the child's peer initiations and peer responses, using the following criteria. Examiners reviewed the observation notes for the 10-minute period. A "pass" (scored as 1) was assigned when interactions (i.e., initiations and responses) were appropriate (i.e., received a + code) and when the

child evidenced at least one complete sentence in the sample. A "fail" (scored as 0) was assigned when peer interactions were generally inappropriate (i.e., 50% or more received a – code), or when the child used only gestures, vocalizations, and/or single words.

The data from the Brigance Inventory of Early Development, parent reports, teacher reports, peer initiations, and peer responses were summed together to yield a total score from 0 to 5 (referred to as the "language ability composite"). Children were identified as TD if their total score was 3 or above. Children were identified as having LLA if their total score was 2 or below. Table 2 displays these results.

Experimental Procedures

The length of the study was 12 weeks. The pretest period was 4 weeks long, followed by two 30-minute mediation sessions over a 4-week period for the MLE group, and no mediation for the control group. Posttesting took place during the final 4-week period. Pretest and posttest data as described below were collected by the first author, who is bilingual and certified in speech-language pathology, and by several bilingual graduate students under direct supervision of the author. The first author and two of the graduate students conducted MLE sessions in groups of two children at the Head Start site in one of two areas provided by Head Start for that purpose. MLE sessions were generally timed to occur during small group time to fit with regular classroom activities. A total of eight MLE sessions (at least two for each mediator) were videotaped for reliability purposes. Children in the no-mediation group participated in regularly planned classroom activities.

Pretest and Posttest Instruments. Three instruments were selected to examine pretest-posttest differences. The instruments were selected to represent a range of similarity and dissimilarity with reports of home language interaction.

TABLE 2. Univariate *F* tests by condition (MLE vs. no mediation) for pretest data.

	MLE		No Mediation			
	M	*SD*	*M*	*SD*	*F*	*p*
Language Ability Composite	3.96	1.41	3.95	1.49	.001	.980
EOWPVT-R	72.25	9.62	73.90	9.92	.430	.514
EOWPVT-RS	6.94	3.47	9.00	4.97	1.531	.228
CSSB	91.65	14.49	91.72	14.94	.000	.986
PLS-A	4.00	2.84	3.83	4.35	.027	.870

Note. Language Ability Composite = the total of Brigance Inventory of Early Development, teacher observation, parent report, classroom observation of peer initiations, and classroom observation of peer responses; 1 point ("pass") contributed by each component measure for a possible total of 5 points. EOWPVT-R1 = Expressive One-Word Picture Vocabulary Test–Revised (*M* = 100, *SD* = 15), pretest. CSSB = Comprehension Subtest of Stanford-Binet Intelligence Scale (*M* = 100, *SD* = 16).; PLS-A = Preschool Language Scale–Adapted (total possible 10 points). EOWPVT-RS = raw score of correct control items on the EOWPVT-R, disregarding 19 items presented during MLE.

Because it is reported that home language experiences in African-American and Latino-American households include focus on descriptions and functions, one measure focused on description. It is further reported that demands for single-word labels are low in these populations. The expectation, based on our clinical experience and on research reports of vocabulary test performance, was that test performance of expressive single-word vocabulary would be low. Thus, single-word vocabulary was the focus of the teaching portion of the dynamic assessment and was a pretest-posttest measure. Finally, a measure that had similar task demands (e.g., vocabulary knowledge) but different content demands (academic language) was selected to examine possible transfer effects of MLE.

Tests were administered in the language(s) of the child using the following guidelines. First, instructions were given in the child's home language. Second, if the child seemed not to comprehend directions or the examples, the examiner switched to the other language for clarification. Assessment then proceeded in the language that was judged to be relatively stronger. Children knew that the examiners were bilingual, and responses given in either language were recorded and scored conceptually (based on Pearson et al., 1993). Third, if children switched between languages during administration of a given test, the examiner switched languages to match the language used by the child. Finally, "no response" answers were noted and, after completion of the test, were administered again in the other language to ensure that the child understood the question.

Description Skills. The Stanford-Binet Test of Intelligence for Children, 4th Ed.: Comprehension Subtest (the pretest and posttest scores are reported as CSSB-1 and CSSB-2, respectively; Thorndike, Hagen, & Sattler, 1986) was used to assess children's use of description. The CSSB is a 42-item test in which children point to body parts, answer questions about the use of common objects (e.g., what do you use an umbrella for?), and respond to process questions (e.g., why shouldn't you run with scissors?). The first seven items involve identification of body parts by pointing to a picture. The rest of the test is verbal with no picture cues. The CSSB was administered and scored according to the test manual. Raw scores and standard scores were calculated for each child.

Single-Word Labeling. The Expressive One-Word Picture Vocabulary Test–Revised (EOWPVT-R; Gardner, 1990) was used to assess performance on single-word labeling and was the focus of the teaching portion of the assessment. In order to account for items that children might know beyond the ceiling, testing was initiated with item 1 and continued at least through item 30. Testing was discontinued when a ceiling was reached after item 30. This procedure ensured a comprehensive measure of vocabulary size and allowed for a sample large enough to estimate change. Because the Spanish version is a translation from English, it is possible that items are not ordered by increasing difficulty in Spanish. Thus, this procedure ensured that pretest-posttest change was not overestimated because a ceiling was obtained early due to item difficulty differences. To compare pretest-posttest performance, the

EOWPVT-R was scored in two different ways. First, the test was scored according to the test manual calculating the basal and ceiling scores (after item 30) to obtain a raw score. This raw score was converted to a standard score using the norms in the test manual. It is important to note that the norming sample is for English speakers. Thus, the reporting of the standard score was done in order to control for maturation. Pretest and posttest standard scores are reported as EOWPVT-R1 and EOWPVT-R2, respectively. Second, a raw score (EOWPVT-RS) was calculated, excluding the 19 items used during MLE from the total. These 19 items were: *bird*, *train*, *duck*, *chicken*, *wheel*, *penguin*, *corn*, *sofa*, *goat*, *tire*, *clothing*, *animals*, *painter*, *food*, *dentist*, *tractor*, *furniture*, *lamb*, and *mail*. Inclusion of test items during mediation allowed us to consider whether pretest-posttest change resulted from presentation of specific items, or whether there was, in fact, carryover of the strategies to additional items. These pretest and posttest raw scores are reported as EOWPVT-RS1 and EOWPVT-RS2, respectively. It is important to note that these items were used during the MLE to illustrate to the child the purpose of using single-word labels but were not items that were "drilled" or "trained" to predetermined criteria. Instead, the items were used as explanations of when to use "special names." For example, in MLE 1 Activity B, which focused on the topic of transportation, emphasis was on the fact that things that take people places have "special names." So, it is possible to say, "I went to see my grandma on the *thing*," but that would be much less clear than using a "special name" such as "*airplane*."

Academic Concepts. The Preschool Language Scale (Zimmerman, Steiner, & Pond, 1978) was used to assess school-related language concepts. Ten items (five expressive and five receptive) in the 3–6 to 4–6 age-range from the Preschool Language Scale were used to measure related language skills. This version (referred to as the PLS-A) was adapted by the first author and included items that were not found on the CSSB or the EOWPVT-R—that is, the receptive items addressed categorization, agent-action, colors, prepositions, and numerical concepts. The expressive items were: *counts blocks*, *names opposites*, *states physical needs*, *counts to 10*, and *comprehends senses*. A summary score representing the total number of correct responses (0–10) was recorded for each child. Pretest and posttest raw scores are reported as PLS-A1 and PLS-A2, respectively.

Mediated Learning Experience. MLE tasks consisted of four activities presented in the child's language(s) in two 30-minute sessions, 1–2 weeks apart. All activities had the general goal of teaching children about labels. This teaching was done through discussion of single-word labels ("special names") in contrast to other ways of referring to objects (e.g., function, location, description, and categorization) and through discussion of the importance of labels. The sessions used toy sets organized by themes to demonstrate how single-word labeling is used. The first MLE session used the themes of transportation and food. The second MLE session consisted of themes about animals and community workers. Materials used during each MLE session included books, cards, puzzles, and toys (see Appendix B for a complete script and materials list). Because a goal of dynamic assessment is to promote self-directed learning, MLE sessions incorporated the strategies of planning and self-regulation as related to children's uses of labels.

A feature of MLE is that the mediator responds to children individually based on their responses and strategies, and to help them modify their learning. Thus, the MLE sessions were not totally scripted; however, the content and materials of the lessons remained constant. A script provided consistency across the activities and ensured that MLE was consistently implemented. For each session, the clinician introduced the activity by stating the goal (mediation of intentionality) and purpose (mediation of meaning):

> Today we're going to play with some toys and books. We're going to think about and say special names. What are we talking about? [elicit child response, "special names"] How do special names help us? They help us know what we're both talking about.

The introduction continued with the clinician relating the planned activity to school and home activities (mediation of transcendence):

> If I call you [wrong name] will you answer me? [elicit "no" response from child] No, because that's not your name. At home you also use special names. What are some special names you use at home? [elicit names of child's family members] What if you call your [child's toy] a [wrong name] instead? Will your [sibling or parent] know what you are talking about? So special names help us know what we mean.

The clinician then helped the children develop a plan for an activity (mediation of competence) and helped them carry out the planned activity.

> So, we're going to play gas station with the cars first. Then we're going to read a book about transportation. What are we doing first? [elicit child response] [Place cars on table without the gas station] Here are some cars, let's play gas station. What do we need? What should we do? [child asks for gas station, in view but out of reach] Do you remember what we are talking about today? [elicit, "special names"] [Continue playing and interacting with child, focusing on special names and their use]

To close the activity, the clinician reviewed the principles of labeling, and described changes observed in the child's ability to label, plan, and self-regulate.

> You worked really hard today. Do you remember what the lesson was about? [elicit "special names"] At first you didn't use special names, but we worked on that and now you know that special names are important. You were able to [name planning strategy specific to child, e.g., look at all the pictures, compare the toys] and [discuss self-regulation strategy specific to the child, e.g., wait for directions, ask when you didn't know the name].

Modifiability Measures. Two scales were used to observe and measure child modifiability: the Learning Strategies Checklist (LSC) (developed by Peña, 1993, based on Lidz, 1991) and the Modifiability Scale (based on Lidz, 1987, 1991). The LSC was used to observe and record children's learning behavior during each of two MLE sessions (LSC-1, LSC-2). The Modifiability Scale was used to summarize the mediator's judgment of overall child change.

Observation of Child Strategies. The LSC (Peña, 1993) was used to record observations of the strategies children used to learn about single-word labels. This instrument consists of 13 three-point items (scored 0–2, for a total of 26 points) that focus on child responsivity (attention, comparative behavior, planning, self-regulation, and motivation) and transfer (near—ability to apply learned strategies within an activity, and far—applying new strategies when a new activity is introduced) during each mediated activity. The checklist consists of three attention/discrimination items (e.g., child initiates a focus with minimal cues from the mediator; child responds to relevant visual cues and ignores irrelevant visual cues), three comparative behavior items (e.g., the child comments on features of the task), two planning items (e.g., the child talks about the overall goal of the session), four self-regulation/self-awareness items (e.g., the child corrects himself or herself), two motivation items (e.g., the child persists even when frustrated), and two transfer items (e.g., the child applies strategies within the same task). These child learning strategies during each of the two MLE sessions were judged for consistency and rated as, "none of the time" (= 0), "some of the time" (= 1), and "most of the time" (= 2). Mediators observed and scored the children's use of these strategies during each of the two MLE sessions, making one rating per item for each session. A total score from 0 to 26 was recorded for each child for each MLE session. These two scores are reported as LSC-1 and LSC-2.

Summary of Change. The Modifiability Scale (based on Lidz, 1987, 1991) was used as a summative observation of child modifiability for the two MLE sessions. This instrument consists of three Likert scale items: examiner effort (4-point scale), child responsivity (4-point scale), and transfer (3-point scale). Examiner effort refers to the amount of support provided to the child during the MLE sessions. Child responsivity refers to the child's flexibility while learning new information. Finally, transfer refers to the child's ability to apply newly learned strategies throughout the MLE sessions. Each of the three Likert scales has brief descriptors for each point (e.g., examiner effort can be scored as extreme, high, medium, and minimal, for 1 to 4 points, respectively). Each child was rated by the mediator, using this instrument, after completing the second MLE session. A sum of the three components yielded a total score for each child from 0 to 11. A high score indicates high child modifiability, whereas a low score indicates low child modifiability.

Mediator Training. Two graduate student mediators were trained on MLE and rating procedures during a 1-week period. Training was conducted using videotaped examples and practice sessions. After discussion of the MLE model, mediators were shown videotaped examples of MLE sessions. Each graduate student rated the mediator's behavior in the examples using the Mediated Learning Experience Rating Scale (MLE-RS; Lidz, 1991, described in the "Fidelity of Treatment" section that follows). Next, the first author developed a teaching script for the mediators to follow during the MLE sessions. Then the first author modeled presentation of the materials, using the script, while the mediators observed. Possible child responses and appropriate mediator reactions within the MLE model were proposed and discussed. Mediators practiced using the materials and script with each other until they were familiar with the procedure.

Mediators were then trained to score the LSC and Modifiability Scales using the same videotaped examples. Mediators independently rated child modifiability using the LSC and Modifiability Scales after viewing the videotape. Scoring differences among the mediators were discussed to reach a consensus. Watching the videotape, scoring, and discussion continued until a criterion of 90% agreement was met, and mediators became familiar with the instruments and scoring procedures.

Fidelity of Treatment. Documentation of consistency of MLE implementation was based on eight (14.8%) sessions that were videotaped at random. Fidelity of treatment was evaluated by scoring the implementation MLE using the MLE-RS (Lidz, 1991; see Appendix A). This scale operationalizes components of MLE that constitute the teaching portion of the dynamic assessment (e.g., intentionality, transcendence, meaning, and competence). Mediator interaction is rated from 0 to 3 for each MLE component. In general, a rating of 3 indicates the highest level of mediation and must include statement of a principle by the mediator to the child. A score of 2 indicates that the component is consistently in evidence and elaborated through explanations or examples. A score of 1 indicates that the MLE component is in evidence, but not elaborated. Finally, a score of 0 indicates that the component was not in evidence. The MLE-RS includes an operational definition and a description of each score (0–3) for each MLE component. Mean MLE scores for the videotaped sessions were calculated as 2.33 (SD = 0.47). This score demonstrates that MLE was consistently implemented during the teaching sessions.

Reliability. Evidence of reliability of the MLE scores was obtained with the help of a bilingual speech-language pathology graduate student not involved in the study. After a 30-minute training session, she reviewed the videotapes and independently scored them using the MLE-RS (Lidz, 1991). Overall point-to-point agreement was 87.5%; agreement within 1 point was 100%.

Reliability was estimated by examining the internal consistency of the scores obtained for the two independent measures of the LSC (LSC-1, based on the first MLE session, and LSC-2, based on the second), using an inter-item correlation matrix. Coefficient alpha was estimated as .82 for the LSC-1 and .85 for the LSC-2. Alpha values between .8 and .9 are considered very good (DeVellis, 1991).

Analyses

This study compared pretest and posttest performance of participants in TD and LLA groups who received MLE, and TD and LLA groups who received no mediation. Descriptive statistics for the TD and LLA groups were calculated for all test tasks, and estimation of pretest-posttest changes was conducted using repeated measures ANOVA. Although there were unequal group sizes in this data set, Cohen (1988) suggested that *F*-tests were generally robust in that even moderate violations of the assumptions of sample size and distribution generally have insignificant validity effects on hypothesis testing. Independent variables included TD and LLA groups; dependent variables included the four measures from three tests (EOWPVT-R, EOWPVT-RS, CSSB, PLS-A). Analyses additionally explored pretest versus posttest differences between MLE and no-mediation groups. Independent variables for these analyses included MLE and no-mediation groups with test scores as the dependent variables. Finally, classification analysis was used to explore how well each measure classified TD and LLA children in the MLE group. Discriminant analysis was also used to determine which combination of the predictor variables best differentiated the TD and LLA groups. These statistical analyses were conducted using the SPSS for Windows 9.0 (1999).

Results

Baseline Measures

Because there was a lack of random assignment, the issue of equality between the MLE and no-mediation groups was addressed first. The assumption was that, because the two groups came from the same community, they represented the same population. Statistical analyses to document this similarity were conducted. Means and standard deviations for the two groups were calculated for descriptive comparisons (see Table 2). One-way analysis of variance (ANOVA) was performed on the baseline data including: language ability composite, EOWPVT-R, EOWPVT-RS, CSSB, and PLS-A. There were no significant differences between the MLE and the no-mediation groups on any of the baseline measures. Possible differences in group composition as a result of L1 versus L2 use, gender, and/or ethnicity were explored using multiple analysis of variance (MANOVA). Fixed factors were identified as language preference (English vs. Spanish), gender, and ethnicity (African-American vs. Latino-American); dependent measures consisted of the four measures taken from the three standardized tests. MANOVA results indicated no significant effects for language, gender, or ethnicity across the language measures. Therefore, it was concluded that the groups were similar in terms of demographic and sociocultural factors.

Differentiation of Language Ability Using Pretest-Posttest Change

The first research question addressed the issue of differences in children's responsiveness to MLE in terms of language ability. This replicated Peña et al.'s (1992) previous research using the EOWPVT (Gardner, 1979) in a pretest-posttest comparison. If dynamic assessment is to be applied as a nonbiased assessment language measure, pretest-posttest differences should be sensitive to true language disorder versus language difference. Thus, pretest versus posttest differences can potentially be used to rule out language disorder. Peña et al.'s (1992) study demonstrated that children with typical development and low language ability were differentiated on their pretest-posttest EOWPVT performance. Thus, we expected to replicate this result with the EOWPVT-R. We further explored whether these effects would carry over to the CSSB at posttest. Because children in the Peña et al. (1992) study with typical development and low language ability demonstrated significantly different mean scores on the CSSB at pretest, we expected that there would be no differential effect from pretest to posttest in the present study. Finally, the PLS-A was added in this study. This measure is somewhat related to the EOWPVT-R with regard to the task demands of using labels. However, it is more academically loaded, and requires knowledge of concepts such as color, location, and part-whole relationships. Therefore, we expected moderate changes in the pretest-posttest performance of the two ability groups.

Four repeated measures ANOVAs were used to compare pretest-posttest performance (EOWPVT-R, EOWPVT-RS, CSSB, PLS-A) of the typically developing and low language ability children who participated in the MLE. For this set of analyses, the between-subjects factor was defined as ability (TD and LLA); the within-subjects factor was defined as time (pretest and posttest), and the dependent variables were defined as the test scores. Results indicated a significant ability difference in the amount of change children made from pretest to posttest on labeling (EOWPVT-R) and academic measures (PLS-A). Table 3 is a display of the pretest and posttest scores for the MLE group. Specifically, for the EOWPVT-R there was a significant main effect for Time [$F(1, 49) = 23.32$, $p < .001$] and Ability [$F(1, 49) = 19.00$, $p < .001$], and a significant Ability × Time interaction [$F(1, 49) = 19.91$, $p < .001$]. For the EOWPVT-RS measure, there was a significant main effect for Time [$F(1, 15) = 7.67$, $p = .014$] and Ability [$F(1, 15) = 13.42$, $p = .002$], and a significant Ability × Time interaction [$F(1, 49) = 7.67$, $p = .014$]. Children identified as typically developing demonstrated greater pretest-posttest changes on the EOWPVT-R and the EOWPVT-RS in comparison to children with low language ability. The EOWPVT-RS result indicates that the change was not attributable to item presentation during the teaching task. Results on the CSSB indicated a main effect for time [$F(1, 45) = 6.21$, $p = .016$] and ability [$F(1, 45) = 9.592$, $p = .003$]. Finally, results on the PLS-A indicated a main effect for Time [$F(1, 45) = 39.49$, $p < .001$] and Ability [$F(1, 45) = 7.834$, $p = .008$].

In order to examine whether pretest-posttest changes resulted from MLE training rather than to 2 months of exposure to the Head Start curriculum, test performance of the children who received MLE was compared to performance of a no-mediation group using repeated measures ANOVA. Furthermore, we wanted to rule out the possible

TABLE 3. Pretest versus posttest means on four measures taken from three standardized tests: TD versus LLA groups.

	Pretest				Posttest			
	TD		LLA		TD		LLA	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
MLE								
EOWPVT-R[a]	72.76	8.08	65.78	4.55	83.56	10.02	66.00	5.60
EOWPVT-RS[b]	7.71	3.29	2.50	0.71	13.13	4.76	2.00	0.00
CSSB[a]	94.05	12.71	78.75	17.47	99.24	13.26	80.00	17.27
PLS-A[c]	4.41	2.74	1.88	2.42	7.10	2.02	3.63	2.83
No Mediation								
EOWPVT-R	75.44	9.39	78.00	7.07	71.41	9.14	75.50	3.54
EOWPVT-RS	9.71	5.28	5.20	2.28	7.38	2.33	6.33	2.52
CSSB	96.88	11.43	75.20	13.39	92.12	12.15	82.50	4.95
PLS-A	5.43	4.53	1.60	3.58	5.13	3.64	5.50	3.54

Note. EOWPVT-R1 = Expressive One-Word Picture Vocabulary Test–Revised, pretest; CSSB = Comprehension Subtest of Stanford-Binet Intelligence Scale; PLS-A = Preschool Language Scale–Adapted; EOWPVT-RS = raw score of correct control items on the EOWPVT-R, with 19 items omitted. Standard deviations appear in parentheses.
[a] Values are standard scores.
[b] Values are number of items correct. Highest score achieved = 26.
[c] Values are number of items correct. Highest possible score = 10.

influence of exposing children to the test twice in a short period of time (practice effect). Comparison of pretest-posttest performance of the MLE group in comparison to the no-mediation group helps to examine these possible practice effects (Cotton, 1998). For this set of analyses, the within-subjects factor was defined as Time (pretest and posttest); the between-subjects factors were defined as Condition (MLE and no-mediation group) and Ability (TD and LLA); the dependent variables were defined as the test scores. Table 3 is a display of the test score means.

Results on the EOWPVT-R showed a significant Time × Condition interaction [$F(1, 65) = 10.59$, $p = .002$] and a Time × Ability × Condition interaction [$F(1, 65) = 4.90$, $p = .03$]. For the control measure, EOWPVT-RS, results indicated a main effect for Ability [$F(1, 64) = 4.26$, $p = .043$, a Time × Condition interaction [$F(1, 64) = 8.34$, $p = .005$] and a Time × Ability × Condition interaction [$F(1, 64) = 5.31$, $p = .024$]. Results on the CSSB demonstrated a main effect for Ability [$F(1, 61) = 6.82$, $p = .011$]. Finally, PLS-A results indicated a main effect for Time [$F(1, 52) = 11.79$, $p = .001$] and a Time × Condition interaction [$F(1, 52) = 5.38$, $p < .024$]. In general, children in the MLE group showed higher pretest-posttest gains compared to children in the no-mediation group, with the exception of the CSSB, which differentiated children by ability at the outset. Furthermore, the EOWPVT-R and EOWPVT-RS measures, which were most closely related to the focus of the MLE, showed differential effects for ability in the MLE and no-mediation conditions. This suggests that the gains made by children with TD and LLA in the MLE condition were not merely the result of accrued experience with the regular Head Start curriculum or test exposure but resulted from the MLE treatment.

Classification Efficacy of Dynamic and Static Measures

The second research question concerned the potential ability of dynamic assessment measures, compared to static measures, to accurately differentiate between language disorder and language difference. We employed classification analysis in order to examine the diagnostic accuracy of each measure and the measures in combination. This procedure entails calculating how well individual children are classified by the language measure, rather than estimating whether children as a group demonstrate significantly different test score means (Gray, Plante, Vance, & Henrichsen, 1999). Plante and Vance (1994) suggested that language tests that discriminate between children with and without language impairment with an accuracy above 90% are considered "good" discriminators and those that discriminate with an accuracy between 80 and 90% are considered "fair" discriminators.

For this set of analyses, we examined the performance of the children who participated in the MLE group. Specifically, the analyses used the dynamic and static measures to predict language ability condition (TD vs. LLA). Posttest scores and modifiability index (MI) (operationalized as the mean of the LSC-1 *Z* scores from MLE Session 1, LSC-2 *Z* scores from MLE Session 2, and Modifiability Scale *Z* scores), were included as the dynamic measures (displayed in Table 4). Note that the *Z* score means for TD children are positive, whereas those for children with LLA are less than –1. The *Z* scores represent the number of standard deviations of the children's scores from the overall group mean. The raw scores were converted to *Z* scores so that each measure would be on the same scale and would have

TABLE 4. Mean modifiability scores: TD and LLA groups.

	TD				LLA			
	Raw Score	*SD*	*Z* Score	*SD*	Raw Score	*SD*	*Z* Score	*SD*
LSC-1	19.91	3.87	.23	.85	15.83	5.22	–1.11	1.05
LSC-2	23.76	2.61	.22	.79	18.75	4.81	–1.24	1.45
MS	7.11	1.02	.27	.75	4.00	1.87	–1.57	0.85
MI	N/A		.20	.63	N/A		–1.28	1.06

Note. TD = typically developing; LLA = low language ability; LSC-1 = Learning Strategies Checklist, 1st MLE session (26 points); LSC-2 = Learning Strategies Checklist, 2nd MLE session (26 points); MS = Modifiability Scale (11 points); MI = Modifiability Index, derived from the mean of LSC-1, LSC-2, and MS *Z* scores.

equal weight. Pretest scores were considered static measures because they represented performance that was independent of MLE. Children's posttest scores were considered dynamic because they represented performance after participation in MLE, hence a score that was indicative of the zone of proximal development.

First, we wanted to explore which combination of predictor variables would yield the best classification accuracy. A stepwise discriminant analysis was performed, with language ability as the dependent variable and the pretest scores, posttest scores, and modifiability scores as predictor variables. This analysis seeks the subset of variables that best characterizes the group differences. Furthermore, the variables are returned in order of maximum effect, with each additional variable entered in the order in which it explains the variance of the function. Results indicate that one function, consisting of the EOWPVT-R2 and the modifiability score, in this order, significantly discriminated between TD and LLA groups [$F(2,42) = 23.54$, $p < .001$]. Classification rates indicated that the function correctly classified 48 out of 52 or 92.3% of the original group cases. The classification rate on the basis of this function demonstrated a specificity rate of 95.3% (classification of TD children as TD), as well as a sensitivity rate for children with LLA of 77.8% (classification of LLA children as LLA).

Next, we wanted to explore the accuracy with which each measure independently classified the participants into TD (with 45 children) and LLA groups (with 10 children). We approached this in two different ways. The first method of classification was based on the mean and standard deviations of the standardized tests and on the means and standard deviations for the adapted standardized measures. This was used to explore what would occur clinically when using these measures for this population. Second, discriminant functions analysis was used to find the optimum classification of each case for each measure. The cut-off score is not given, as the approach uses probability to classify each case, but it can be inferred from the classification.

For the first classification analysis, the cut-off scores were set at one standard deviation below the mean for the EOWPVT-R and the CSSB. In addition, the cut-off scores for the adapted standardized measures and modifiability measures were set at one standard deviation below the mean for the typically developing group at pretest. Table 5 displays the classification rates using the one standard deviation cut-off score for each measure. Overall, static (pretest) measures showed high degrees of misclassification, ranging from 25% overall correct classification for the EOWPVT-R1 to 90% overall correct classification for the PLS-A1. The EOWPVT-R1 classified all the children in the TD group as LLA, whereas the PLS-A1 classified 50% of children with LLA as TD. The EOWPVT-RS1 demonstrated fair sensitivity and specificity rates; however, this is likely a result of the lower raw scores of the LLA group after deletion of items presented in the MLE session. The CSSB-1 also demonstrated fair sensitivity and specificity rates; 12% of the children in the LLA group were incorrectly classified as TD, and 21% of the children in the TD group were incorrectly classified as LLA. Classification improved for the posttest score measures with the exception of the PLS-A2. MI and the EOWPVT-RS2 had the best overall classification rates of 92%; however the EOWPVT-RS2 demonstrated a poor sensitivity rate of 56%.

The second set of analyses was conducted using a discriminant functions analysis, which classified cases based on a distance function. The program computes the probability of each case's fit with each group based on the distance from the mean. Each case is then classified based on highest probability of fit. Cut-off score ranges can be estimated based on the highest score among cases classified as LLA and the lowest score among cases classified as TD. Table 6 displays the sensitivity, specificity, and overall classification rates for these analyses. Consistent with the previous analyses, the pretest measures showed a higher error rate than the posttest measures, ranging from 83% for the EOWPVT-R1 to 88% for the CSSB-1. The pretest measures demonstrated poor sensitivity, ranging from 0% identification of children with LLA as LLA for the EOWPVT-R1, EOWPVT-RS1, and PLS-A1 to 25% for the CSSB-1. However, specificity rates were very good, classifying children with TD as TD 100% of the time. On the other hand, posttest measures demonstrated higher overall classification rates, ranging from 89% for the EOPWVT-R2 to 94% for the EOWPVT-RS2 and PLS-A2. The EOWPVT-R2 demonstrated fair to good sensitivity and specificity, with slightly lower sensitivity, but higher specificity for the EOWPVT-RS2. The CSSB-2 and

TABLE 5. Classification accuracy for individual measures using 1 *SD* as the cut-off score.

	Cut-Off Score	Sensitivity [d]		Error (LLA as TD)		Specificity [e]		Error (TD as LLA)		Overall Percentage of Correct Classification	Projected Classification Assuming Equal Size TD and LLA Groups [f]
		Proportion	%	Proportion	%	Proportion	%	Proportion	%		
EOWPVT-R1 [a]	85	9/9	100	0/9	0	4/43	9	39/43	91	25	55
EOWPVT-R2 [b]	85	10/10	100	0/10	0	26/42	68	17/42	32	69	84
EOWPVT-RS1 [a]	4.07 [c]	7/8	88	1/8	22	35/43	81	8/35	19	82	85
EOWPVT-RS2 [b]	4.07 [c]	5/9	56	4/9	44	44/44	100	0/44	0	92	78
CSSB1 [a]	84	7/8	88	1/8	12	34/43	79	9/43	21	80	84
CSSB2 [b]	84	6/8	75	2/8	25	37/41	90	4/41	10	88	83
PLS-A1 [a]	1.57 [c]	4/8	50	4/8	50	43/44	98	1/44	2	90	74
PLS-A2 [b]	1.57 [c]	2/8	25	6/8	75	40/40	100	0/40	0	88	63
MI [b]	–1 [c]	7/9	78	2/9	22	42/44	95	2/44	5	92	87

Note. EOWPVT-R = Expressive One-Word Picture Vocabulary Test–Revised, EOWPVT-RS1 = EOWPVT-R Raw Score Control (raw score less the 19 items presented during MLE); CSSB = Comprehension Subtest of the Stanford Binet Test of Intelligence for Children; PLS-A = Preschool Language Scale, 10-Item Adaptation; MI = Modifiability Index (derived from the mean of LSC-1, LSC-2, and MS *Z* scores); 1 and 2 = pretest and posttest, respectively. Slight differences in the number of possible TD and LLA cases are the result of missing data for specific analysis.

[a] A static measure.
[b] A dynamic measure.
[c] Cut-off scores are 1 standard deviation below the TD group mean at pretest.
[d] Sensitivity = correct classification of children with LLA as LLA.
[e] Specificity = correct classification of children with TD as TD.
[f] Illustrates projected overall classification with equally weighted groups, obtained by averaging specificity and sensitivity together, based on current statistics.

TABLE 6. Discriminant analysis for individual measures using the optimum cut-off score.

	Cut-Off Score	Sensitivity [d]		Error (LLA as TD)		Specificity [e]		Error (TD as LLA)		Overall Percentage of Correct Classification	Projected Classification Assuming Equal Size TD and LLA Groups [f]
		Proportion	%	Proportion	%	Proportion	%	Proportion	%		
EOWPVT-R1 [a]	<55 [c,f]	0/9	0	9/9	100	43/43	100	0/43	0	83	50
EOWPVT-R2 [b]	70 [d]	8/10	80	2/10	20	40/44	91	4/44	9	89	86
EOWPVT-RS1 [a]	<2 [c]	0/8	0	8/8	100	43/43	100	0/43	0	84	50
EOWPVT-RS2 [b]	5–6 [e]	7/9	78	2/9	22	43/44	98	1/44	2	94	88
CSSB1 [a]	66–70 [e]	2/8	25	6/8	75	43/43	100	0/43	0	88	63
CSSB2 [b]	70–74 [e]	3/8	38	5/8	62	41/41	100	0/41	0	90	69
PLS-A1 [a]	0 [c,f]	0/8	0	8/8	100	44/44	100	0/44	0	85	50
PLS-A2 [b]	3–4 [e]	5/8	63	3/8	37	40/40	100	0/40	0	94	82
MI [b]	–1.03 to –1.16 [e]	6/9	67	3/9	33	42/44	96	2/44	4	91	82

Note. EOWPVT-R = Expressive One-Word Picture Vocabulary Test–Revised, EOWPVT-RS1 = EOWPVT-R Raw Score Control (raw score less the 19 items presented during MLE); CSSB = Comprehension Subtest of the Stanford Binet Test of Intelligence for Children; PLS-A = Preschool Language Scale, 10-Item Adaptation; MI = Modifiability Index (derived from the mean of LSC-1, LSC-2, and MS Z-scores); 1 and 2 = pretest and posttest, respectively. Slight differences in number of possible TD and LLA cases are the result of missing data for specific analysis.

[a] A static measure.
[b] A dynamic measure.
[c] All cases classified as TD; cut-off score is the lowest score achieved.
[d] Score is at boundary. Cases scored at boundary have an equal probability of being classified as TD or LLA.
[e] Low score represents highest score achieved by a case classified as LLA. High score represents lowest score achieved by a case classified as TD. Cut-off score is in this range.
[f] Cut-off score is the lowest possible score.
[g] Sensitivity = correct classification of children with LLA as LLA.
[h] Specificity = correct classification of children with TD as TD.
[i] Illustrates projected overall classification with equally weighted groups, obtained by averaging specificity and sensitivity together, based on current statistics.

PLS-A2 sensitivity rates were poor (38% and 63% respectively), with good specificity rates (100%). The MI had an overall classification of 91%, with poor sensitivity (67%) and very good specificity (96%).

Comparison of classification accuracy indicated that, regardless of analysis approach used, no single measure reached the 90% accuracy level. Several measures reached the 80% accuracy criterion. As predicted, the majority of the dynamic measures had a classification accuracy greater than 80%, whereas static measures (pre-MLE) fell below this level.

When comparing across classification approaches, the discriminant approach seems superior with respect to overall classification. This difference should be noted with caution because discriminant approaches do not work well when group sizes are disproportionate, as in this study, because high error rates of the smaller group can be masked by low error rates of the larger group. When the LLA and TD groups were equally weighted, and sensitivity and specificity rates averaged, there was no clear advantage of one analysis procedure over the other with respect to overall correct classification. The two methods, however, do result in important differences with respect to the specificity and sensitivity of the measures. The specificity of all measures increases, or remains high, with the use of discriminant analysis. On the other hand, sensitivity tends to decrease with the use of discriminant analysis, with this decrease being most evident and substantial for the nondynamic measures.

Discussion

This study used a test-teach-retest paradigm to: (a) explore whether children with typical development and those with low language ability respond differently to short term MLE, and (b) compare the discriminant validity of dynamic and static assessment measures in differentiating language difference and language disorder.

Language Ability Differences

When comparing children who are typically developing with those who have low language ability, it was seen that the typical children benefited the most from short-term MLE within a dynamic assessment framework. By giving the children experience with the expectations of the test task through MLE, the importance of single-word labeling was emphasized. Children who had low pretest scores, but normal ability, were able to profit significantly from the MLE on assessment measures that required labeling, even though MLE had a brief duration. These children made significantly greater gains from pretest to posttest than children with LLA and children in the no-mediation group on both the EOWPVT-R and the EOWPVT-RS, as reflected by the Time × Ability × Condition interaction in the repeated measures ANOVA. Children in the MLE condition made greater gains on the PLS-A in comparison to children in the no-mediation condition, as reflected by the Time × Condition interaction in the repeated measures ANOVA, but children in the TD and LLA groups did not respond differentially to MLE. This is likely due to significant differences between the two ability groups at the outset on this measure. The CSSB did not demonstrate interactions for ability or condition; however, this is to be expected because the pretest alone already significantly differentiated the two ability groups.

These results have implications for assessment of nonmainstream children who frequently do not perform well on standardized tests and may be misidentified as having language disabilities. Budoff (1987) argued that training-based assessment of children from nonmainstream cultures allowed them to perform optimally once they understood the task demands of the test. Children who have normal language-learning ability but limited experience are expected to learn new language rules readily. On the other hand, children who have true language-learning problems are expected to have difficulty learning new language rules.

These findings are consistent with dynamic assessment studies that found that with MLE, children from nonmainstream groups are able to improve their standardized test scores (Carlson & Wiedl, 1980; Feuerstein, Miller, Rand, & Jensen, 1981; Lidz & Macrine, 2001; Lidz & Thomas, 1987; Missiuna & Samuels, 1989). This higher score, or ability to change, is typically considered to be a more valid measure of ability. The MLE helps to rule out situational bias by providing children with the "rules of the game."

An issue in using a pretest-teach-posttest approach is that exposure to the test itself may account for increased scores. Our results comparing the MLE and the no-mediation groups indicated that test exposure alone could not explain the pretest-posttest changes of the MLE group. For three of four measures, there was a significant Condition × Time interaction favoring the MLE condition. Another possibility is that increased individual time with children and with materials may explain the pretest-posttest changes. We do not believe this is likely because the interventions were timed to occur during classroom-based small group activities. During each school day, all children rotated through 20- to 30-minute periods of work with the teachers in small groups of up to 4 children each. While some children worked in these groups, others in the classroom were involved in self-directed activities. Therefore, all children were exposed to small group instruction and the increased attention from the teacher that it allowed. Additionally, we have explored this possibility by designing a study that examines differential effects of teaching techniques within a dynamic assessment framework (Stubbe-Kester, Peña, & Gillam, in review). However, we cannot entirely rule out the effects of such special attention.

To control for the possibility that exposure to actual test items accounted for test gains, we examined pretest-posttest changes on the EOWPVT-R in two ways. First, we looked at the standard score differences that included the 19 items from the test that were included in the MLE, and second, we compared the raw score differences, excluding any of these items that the children correctly identified when tested. Pretest-posttest analyses indicated the same performance patterns for both scoring methods. Children

who are typically developing increased their correct responses to items that were not part of the MLE, whereas children who were identified as having low language ability did not demonstrate significant change on either measure. Thus, the ability to transfer what was learned to novel items seemed to be an important element in differentiating the two groups.

There are two possible explanations for the role that MLE played in raising children's test scores. It is possible that the children already had the items in their repertoire but did not understand the pretest task. Thus, MLE would have helped children understand the test task. Another possibility is that after MLE, they became more aware of labels and the need to use labeling in their daily activities. This awareness helped them to learn new single-word labels in addition to those presented during the MLE. Further, there is the issue of comfort and rapport in the assessment situation. The nature of dynamic assessment may be a better fit to the nature of adult-child interactions within the home. These possible explanations are yet to be tested.

The literature on dynamic assessment suggests that when children learn linguistic strategies, they are able to transfer mediated skills across tasks. One finding in our study that supports this notion is that TD children in the MLE group improved their performance on all the language measures, even though MLE focused on single-word labeling of objects. For example, the PLS-A tested knowledge of language skills related to academic language (e.g., colors, opposites, and prepositions) that require precision. Thus, children appeared to become more aware of the importance of referential specificity in academic language. With respect to improvement in the area of description on the CSSB task for the MLE group, it may be that children learned to be more specific during description tasks in addition to labeling tasks, which resulted in higher CSSB posttest scores. For example, one question on the CSSB is, "Why do people use umbrellas?" In the pretest children typically responded, "For the rain," which, according to the scoring rules, is scored as 0. In the posttest, more children responded, "To keep from getting wet in the rain." (scored as 1), to this same question. Thus, provision of more specific information was the difference between a correct and incorrect score. This evidence is consistent with the idea of the "snowball effect" of mediation enabling children to become active learners (Haywood & Wingenfeld, 1992; Lidz, 1996).

Predictive Validity of Dynamic Assessment and Static Measures

Classification analysis of the individual measures generally indicated higher misclassification for the static (pretest) measures in comparison to the dynamic measures. Additionally, readers are cautioned that the use of static tests may result in unacceptably high false positive rates (classification of children with TD as LLA), particularly for children from African American and Latino American backgrounds. Although children from TD and LLA groups significantly differed on the "static" CSSB and PLS-A, classification analysis of those measures underscores the notion that a significant difference may not have a large enough effect for accurate classification. For example, although the performance difference of children with typical development and low language ability on the CSSB-1 was statistically significant, overall classification accuracy using a discriminant approach was 88% with a 25% sensitivity rate and a 100% specificity rate, resulting in a high number of false negatives. However, from a clinical point of view, we are not as interested in overall classification rates as in accurate identification of children with low language ability. Using a 1 *SD* cut-off approach, the overall classification rate was slightly lower, at 80%. However, the sensitivity rate was higher, at 88%, and the specificity rate lower, at 79%. This approach would then yield a high false positive rate. Using either approach, this measure alone could not be used with confidence to distinguish between children with TD and LLA.

These results emphasize the risk of over-identifying children with TD as LLA if relying on normed tests that are not appropriate for the population. The discriminant approach results emphasize that sensitivity could be lost—even with "optimal" cut points—if the test is not designed for the population in the first place.

The main thrust of this research was to examine whether dynamic assessment could be used to reduce the high rate of nonmainstream children identified as language disordered. The fact that EOWPVT-R2 and MI together provide the most accurate classification of the TD and LLA groups suggests that dynamic assessment adds information not evident in static assessment (i.e., pretest scores). Examination of the discriminant accuracy indicated that these measures taken together were fair to good discriminators of TD and LLA children. Although sensitivity of this function (78%) was below expectations (i.e., 90%), specificity was 95%. Furthermore, analysis of the 4 misclassified cases indicated that the composite language ability scores used to initially classify the children were near the boundaries of the TD and LLA cut-off scores. One factor that may have influenced this result is that children in the participant pool had varying language abilities that ranged from low to high. Thus, it is reasonable to expect that some children would not clearly fall into one of the two ability groups. Furthermore, these results highlight the limitations of the discriminant approach, in that it is highly population dependent. That is, optimal classification is determined for a particular data set. With more equal-sized groups and with different populations (hence different scores), such an approach would likely yield different cut-off scores.

The findings of this study have clinical implications for differentiating language difference and language disorder. The dynamic assessment measures predict that a highly "modifiable child" (i.e., one who received a high score on the Modifiability Scale) who shows test gains in the area targeted for teaching (in this case single-word labels), is one who is likely to have normal language learning abilities. On the other hand, a child who demonstrates limited modifiability and little pretest versus posttest change after short-term MLE is predicted to have limited language abilities. In this study, children who were highly

modifiable were able to initiate and maintain focus, plan, self-regulate, transfer strategies, and maintain motivation. Additionally, these children were highly responsive during intervention, required little examiner effort to change during the intervention, and showed evidence of transfer during interventions and posttesting. During the interventions, the more modifiable children readily played with the materials, were able to respond to questions regarding the purpose of the tasks, and showed good problem-solving skills. When the task became difficult, they tried to generate problem-solving strategies or looked to their peers and/or the mediator for help and support. On the other hand, children who scored low on modifiability measures needed much redirecting, encouragement, and praise to persist in the task. These children became frustrated if the task became too difficult and tended to want to quit. Their posttest scores tended to show limited change. Analyses of children's learning strategies and responsiveness to MLE are valuable additions to current assessment procedures because they can provide information about how the child learns, as well as the learning process that may need to be targeted in intervention. Thus, it is necessary to determine both the zones of actual and proximal development in order to more fully understand the level of functioning of the child.

Overall, these results are consistent with Vygotsky's notion of the zone of proximal development. It was clearly seen that children who had similar initial scores on the EOWPVT-R showed different learning profiles when given mediated learning. These individual learning profiles help the clinician to make a determination of language disorder versus language difference. For children with low language ability, individual modifiability scores potentially guide intervention decisions (Bain & Olswang, 1995).

Thus, the application of dynamic assessment to language assessment is promising for children from diverse language backgrounds. Based on the results of this study, future research should focus on modifiability and the long-term effects of MLE. First, studies further operationalizing and quantifying modifiability are needed. Modifiability seems to be a valid indicator of language ability but needs further profiling for specific behaviors and subtle changes that occur over several sessions. Additionally, modifiability in children with language impairment should be further investigated. Of particular clinical interest is targeting appropriate intervention goals based on both initial dynamic assessment and ongoing observation of child modifiability. The benefits of dynamic assessment for children from nonmainstream backgrounds and for children with true language disorders warrant continued investigation.

Acknowledgments

This research was supported in part by a grant from the National Institutes of Health grant (NIDCD) DC0014102, awarded to Dr. Elizabeth Peña. The authors would like to thank Rena Krakow and Vera Gutierrez-Clellen for their thoughtful comments on earlier drafts of this work. We gratefully acknowledge the cooperation and participation of the Head Start Program administrators, teachers, and children at the various centers, who provided us with valuable experiences and information. In addition, we acknowledge the contributions of the following individuals in collecting data: Maggie Campbell, Kristin Youngdahl, Michele Aurignac, Pnina Siegler, Maria Maunez, Janet Quiñones, and Brian Goldstein.

References

Anderson-Yockel, J., & Haynes, W. O. (1994). Joint book-reading strategies in working-class African American and White mother-toddler dyads. *Journal of Speech and Hearing Research*, *32*, 583–593.

Bain, B. A., & Dollaghan, C. A. (1991). The notion of clinically significant change. *Language, Speech, and Hearing Services in Schools*, *22*, 264–270.

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.

Bates, E., Bretherton, I., & Snyder, L. (1991). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, England: Cambridge University Press.

Bereiter, C. (1985). The changing face of educational disadvantagement. *Phi Delta Kappa*, *66*, 538–541.

Blount, B. (1982). Culture and the language of socialization: Parental speech. In D. A. Wagner & H. W. Stevenson (Eds.), *Cultural perspectives on child development* (pp. 55–76). San Francisco, CA: H.W. Freeman.

Brigance, A. (1978). *Brigance Diagnostic Inventory of Early Development.* North Billerica, MA: Curriculum Associates.

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 Press.

Carlson, J. S., & Wiedl, K. H. (1980). Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical formulations. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *1*, 303–318.

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

Cotton, J. W. (1998). *Analyzing within-subjects experiments.* Mahwah, NJ: Lawrence Erlbaum.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy.* Austin, TX: Pro-Ed.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Education Review*, *56*, 18–36.

Dale, P., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, *16*, 239–249.

Damico, J. S. (1990). Prescriptionism as a motivating mechanism: An ethnographic study in the public schools. *Journal of Childhood Communication Disorders*, *13*, 85–92.

Della Corte, M., Benedict, H., & Klein, D. (1983). The relationship of pragmatic dimensions of mothers' speech to the referential-expressive distinction. *Journal of Child Language*, *10*, 35–43.

DeVellis, R. (1991). *Scale development: Theory and applications* (Vol. 26). Newbury Park, CA: SAGE Publications.

Duran, R. (1989). Assessment and instruction of at-risk Hispanic students. *Teaching Exceptional Children*, *56*, 145–152.

Ellis Weismer, S., & Hesketh, L. (1993). Hypothesis-testing abilities of language-impaired children. *Journal of Speech and Hearing Research*, *34*, 1013–1025.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques.* Baltimore: University Park Press.

Feuerstein, R., Miller, R., Rand, Y., & Jensen, M. (1981). Can

evolving techniques better measure cognitive change? *Journal of Special Education, 15,* 201–219.
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.
Gersten, R., & Woodward, J. (1994). The language-minority student and special education: Issues, trends, and paradoxes. *Exceptional Children, 60,* 310–322.
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.
Gutierrez-Clellen, V., & Iglesias, A. (1987, November). *Expressive vocabulary of kindergarten and first grade Hispanic students.* Paper presented at the American Speech-Language-Hearing Association, New Orleans, LA.
Gutierrez-Clellen, V., Peña, E., & Quinn, R. (1995). Accommodating cultural differences in narrative style: A multicultural perspective. *Topics in Language Disorders, 15*(4), 54–67.
Gutierrez-Clellen, V., & Quinn, R. (1993). Assessing narratives in diverse cultural/linguistic populations: Clinical implications. *Language, Speech, and Hearing Services in Schools, 24,* 2–9.
Hamayan, E., & Damico, J. (1990). *Limiting bias in the assessment of bilingual students.* Austin, TX.: Pro-Ed.
Haywood, H. C., & Wingenfeld, S. A. (1992). Interactive assessment as a research tool. *Journal of Special Education, 26,* 253–268.
Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* Cambridge, England: Cambridge University Press.
Heath, S. B. (1986). Sociocultural contexts of language development. In C. Cortes (Ed.), *Beyond language: Social and cultural factors in schooling language minority students.* Los Angeles: Evaluation and Dissemination Center, California State University.
Heller, K., Holtzman, W., & Messick, S. (1982). *Placing children in special education: A strategy for equity.* Washington DC: National Academy of Sciences.
Iglesias, A. (1985). Cultural conflict in the classroom: The communicatively different child. In D. N. Ripich & F. M. Spinelli (Eds.), *School discourse problems* (pp. 79–96). San Diego, CA: College-Hill.
Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language, 20,* 523–549.
Kiernan, B., & Gray, S. (1998). Word learning in a supported learning context by preschool children with SLI. *Journal of Speech, Language, and Hearing Research, 41,* 161–171.
Kovarsky, D., & Maxwell, M. M. (1992). Ethnography and the clinical setting: Communicative expectancies in clinical discourse. *Topics in Language Disorders, 12*(3), 76–84.
Lidz, C. (1996). Dynamic assessment and the legacy of L. S. Vygotsky. *School Psychology International, 16,* 143–154.
Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to evaluating learning potential.* New York: Guilford Press.
Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment.* New York: Guilford Press.
Lidz, C. S. (1995). Dynamic assessment and the legacy of L. S. Vygotsky. *School Psychology International, 16,* 143–153.
Lidz, C. S., & Macrine, S. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners. *School Psychology International, 22,* 74–96.
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* (p. 511). New York: Guilford Press.
Minick, N. (1987). Implications of Vygotsky's theories for dynamic assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 116–140). New York: Guilford.
Missiuna, C., & Samuels, M. (1989). Dynamic assessment of preschool children with special needs: Comparison of mediation and instruction. *Remedial and Special Education, 10*(2), 53–59.
Parnell, C. F. (1994). *A comparison of screening instruments with a multicultural population: Direct screening versus parent report and teacher observation instruments.* Unpublished master's thesis, University of Texas, Austin.
Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43,* 93–120.
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. (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.
Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15–24.
Quinn, R. (1992). *Mother-infant interactions among Puerto Rican dyads.* Unpublished doctoral dissertation, Temple University, Philadelphia, PA.
Reinharth, B. M. (1989). *Cognitive modifiability in developmentally delayed children.* Unpublished doctoral dissertation, Yeshiva University, New York, NY.
Rogoff, B. (1991). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.
Schieffelin, B. B., & Ochs, E. (1986). *Language socialization across cultures.* Cambridge, England: Cambridge University Press.
SPSS. (1999). SPSS for Windows (Version 9.0) [CD-ROM]. Chicago: Author.
Stubbe Kester, E, Peña, E. D., & Gillam, R. B. (in review). *Outcomes of dynamic assessment with culturally and linguistically diverse students: A comparison of three teaching methods.* Manuscript submitted for review.
Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.
Tomasello, M., & Farrar, M. (1986). Joint attention and early language. *Child Development, 57,* 1454–1463.
Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language, 4,* 197–211.
Vincent, L. J. (1992). Families and early intervention: Diversity and competence. *Journal of Early Intervention, 16,* 166–172.
Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA: MIT Press.

Westby, C. E., & Erickson, J. (1992). Prologue. *Topics in Language Disorders*, *12*(3), v–viii.
Wright, P., & Santa Cruz, R. (1983). Ethnic composition of special education programs in California. *Learning Disability Quarterly*, *6*, 387–394.
Yates, J. R. (1988). Demography as it affects special education. In A. Ortiz & B. Ramirez (Eds.), *Schools and the culturally diverse: Promising practices and future directions.* Reston, VA: Council for Exceptional Children.
Zimmerman, I., Steiner, V., & Pond, R. (1978). *Preschool Language Scale.* Columbus, OH: Charles E. Merrill.

Received October 16, 2000
Accepted January 19, 2001
DOI: 10.1044/1058-0360(2001/014)

Contact author: Elizabeth Peña, PhD, Department of Communication Sciences and Disorders, 2504 Whitis, Austin, TX 78705.

Appendix A

Selected MLE Components

MEDIATED LEARNING EXPERIENCE (MLE) RATING SCALE
(for use with parent-child, teacher-child, examiner-child interactions with preschool children)

Developed by Carol S. Lidz, PsyD
Based on the theory and research of Prof. Reuven Feuerstein

Child: _____ Mediator: _____ Task: _____

Rater: _____ Date: _____ Location: _____

INTENTIONALITY: a conscious attempt by the mediator to influence the behavior of the child. This includes communication to the child of the purpose for the interaction, as well as attempts by the mediator to maintain the child's involvement in the interaction. For children who are already self-regulating and do not require interventions by the mediator to engage them in the activity, rating of intentionality includes the readiness of the mediator to become involved as necessary; therefore, the mediator shows ongoing interest in the activity involvement of the child (in this case, the rating would be a 2, unless a statement of a principle is provided).

0 = no evidence
1 = inconsistently present; loses involvement
2 = consistently in evidence
3 = in evidence, with statement or encouragement of a principle to induce self-regulation in the child; this principle would apply to the child's ability to maintain attention and inhibit impulsivity
NOTES:

TRANSCENDENCE: promotion of cognitive bridges between the task or activity and related but not currently present experiences of the child; these may refer to the past or may anticipate the future. These bridges must promote visual images and help to move the child from the perceptual to the conceptual.

0 = not in evidence
1 = simple, nonelaborated reference to past or future experience
2 = elaborated reference
3 = elaborated reference that includes hypothetical, inferential, or cause-and-effect thinking
NOTES:

MEANING: moving the content from neutral to a position of value and importance; this may be done by affective emphasis or stating that the object or aspect of focus is important and should be noticed (or, in contrast, that it is negative and to be ignored or avoided).

0 = not in evidence
1 = calling up labels or concepts already within the child's repertory; saying that it is important and should be noticed (e.g., "Look at this"), but without elaboration
2 = adding animation or affect to make the activity come alive and provoke interest
3 = elaboration that expands the information about the activity or object; this elaboration addresses information that is perceptible to the child within the situation
NOTES:

COMPETENCE (Task Regulation): manipulation of the task to facilitate mastery by the child.

0 = not in evidence
1 = simple directions or passive manipulation of the task (e.g., holding it, moving pieces toward the child, building a model without elaborated directions)
2 = elaborated directions; nonverbal organization into a kind of conceptual grouping
3 = induction/statement/encouragement of strategic thinking and a planful attitude (e.g., "Where shall we start?" "What should we do first?"), or statement of a principle that the child can use to solve similar problems

NOTES:

Appendix B

MLE SCRIPT and Materials

Introduction for Each Mediation

[f]Today we're going to play with food and transportation toys [varied according to the activities]. [f,a]We're going to think about and say special names. [b]How do special names help us? [they help us tell things apart] [b]If I call you hey you will you know who I mean? [no] [c]At home you also use special names. [c]What are some special names you use for family members? [help child generate family member's names] [c]What if you call Susan, Terry, instead? Will your sister (mommy, brother, daddy, dog, etc.) know who (what) you are talking to (about)? [a]Special names help us tell things apart.

Mediation 1

Activity A

Materials: food puzzle, plastic refrigerator with plastic food and containers.

Opening: [a]So we're going to do a puzzle about food, and then play with a refrigerator. [a]First we're going to complete this puzzle. [e]Let's look at it. All these things are food. Do you know the special names? [name them with child] How do we do the puzzle? [c]If I turn it over this way can we do it? [no] [a]What do we need to do? [turn it over, take out the pieces]

Activity: Now let's play with the refrigerator. [b]Can you think of some things with special names that go inside a refrigerator? [name food] There are some things inside here. [a]Do you know what they are? [no] [a]What do we need to do to find out what's inside? [open the door] [a]Let's name these. [a]Do they all go inside the refrigerator? [no] [c]Why not? [they're not all food]

Closing: [f]What did we do? [a puzzle about food, play with a refrigerator] [b]What did we think about? [special names] [g]You did very well, at first you didn't know the names, but when we thought about it you were able to tell me the names. [depends on specific child change.]

Activity B

Materials: book—"Airplanes and things that fly"; toys—truck, car, fire engine, jeep, truck.

Opening: [a]Now we're going to read a book about things that fly, and play with some cars. [b]Do you remember what we're thinking about today? [special names] [c]Let's read the book. How do we do this? [a]Can I open the book in the middle? [no, you must start at the beginning] [name and talk about the pictures]

Activity: [a]Now we're going to play with some cars. [a]Let's look at the track first. [a]Where can the cars go? [around, in the tunnel, down the ramp] [b]What are some special names of cars? [jeep, truck, fire engine, police car, etc.] [b]They have special names and special jobs.

Closing: (same as above)

Mediation 2

Activity C

Materials: Animals—book "Baby Animals", puzzle, Li'l Playmates Farm.

Introduction (above):

Opening: [f]Today we're going to look at different animals. [a]We're going to read a book, do a puzzle, and play with a farm. [b]Do you remember what we're going to think about and say? [special names]. [a]Let's read the book first. [a]What do I have to do to read the book? [open it, turn the pages] [name the animals]

Activity: [a]Now let's complete the puzzle. [a]Do you know these special names? [name them] [a]They are all animals and they also have special names. [c]If I made these pieces very big would they still fit? [no] [c]What if I made them smaller? [there would be too much space] [name the puzzle pieces and complete the puzzle] [a]Now let's play with the farm. [a]How are you going to set this up? [many possibilities, fence around, animals inside, some outside] [b]What are we thinking about today? [special names] [go on to name the pieces of the farm equipment and the animals]

Closing: (above)

Activity D

Materials: Card game about jobs, book about rooms in a house.

Opening: [a]Now we're going to play a card game and read a book. [a]First let's play with these cards. We're going to match these cards. [b]Do you remember what we're talking about? [special names] [b]These people have special names too. [a]Let's look at the cards and find where they go. [name workers] [c]What do they do? [talk about jobs] [b,c]They have special names and special jobs. [c]Have you ever seen a ballerina?

Activity: [a]Let's look at the things in this book. [a]Do you remember how to read a book? [a]Can I start in the middle? [a]Turn it upside down? [no, have to start at the beginning] [c]These are all things found at home. [a]What are the special names? [name by group; furniture, tools, clothes, toys, kitchen equipment] [c]Do you have a refrigerator at home? [c]Does it look like this or is it different? [b]Each thing has a different use and a different name. Closing: (above)

Note. a = mediation of competence (task regulation); b = mediation of meaning; c = transcendence; d = mediation of change; e = sharing; f = intentionality.