

THE BILINGUAL RESEARCH JOURNAL
Winter 1996, Vol. 20. No. 1, pp. 93-131

THEORETICAL AND PRACTICAL IMPLICATIONS OF
ASSESSING COGNITIVE AND LANGUAGE DEVELOPMENT
IN BILINGUAL CHILDREN WITH QUALITATIVE METHODS

Virginia Gonzalez
Patricia Bauerle & Maria Felix-Holt
University of Arizona

Abstract

This study has the objective of using qualitative data to support theoretical and practical implications of important methodological problems affecting the assessment of bilingual children's cognitive and language development. Three instruments were used for accurately identifying gifted students among seventeen Hispanic bilingual kindergartners (first, second, and third generation Mexican-Americans) from low socioeconomic backgrounds. (1) a Home Language Survey, (2) a locally-designed Teachers' and Parents' Rating Scale of Creativity, and (3) the Qualitative Use in English and Spanish of Tasks (QUEST) measuring cognitive and language development in bilinguals (Gonzalez, 1991, 1994, 1995). Using chi-square tests and case studies, six interacting patterns were found indicating the influence of first and second language, verbal and non-verbal assessment procedures, multiple measurements and informants, individualizing assessment, and evaluators' personalities on the assessment of bilingual children's cognitive and language development.

This article examines the assessment of cognitive and language development in bilingual children with a twofold purpose: (1) at the theoretical level, we aim to critically review some of the most important methodological problems, and to derive some theoretical and practical

implications from a qualitative analysis perspective; and (2) at the applied level, we aim to accurately identify gifted bilingual Hispanic kindergarteners employing the Qualitative Use of English and Spanish Tasks (QUEST) developed by Gonzalez (1991, 1994, 1995). This qualitative assessment method measures the interface between cognition, culture, and language in bilingual children; and it can accurately differentiate second-language learning from genuine giftedness.

Since currently we are still dealing with methodological problems when assessing bilingual children, a large metropolitan school district in the Southwest, with more than 50% language-minority children, adopted QUEST experimentally. This instrument was adopted for overcoming the underrepresentation of Hispanic bilingual children in gifted education programs, generating discrimination problems with legal and sociopolitical implications in the community. This adoption was successful as supported by data presented in this study from 17 bilingual children who were evaluated using QUEST.

To accomplish this twofold purpose, a critical literature review of major methodological problems affecting the assessment of bilingual children, and qualitative analysis of results using chi-square tests and multiple and single case studies were used for generating six patterns that will be illustrated using excerpts (for a complete case study analysis of the implementation of QUEST, see Gonzalez, Bauerle, and Felix-Holt, 1994a). In the discussion of these six patterns emphasis will be given to theoretical and applied implications when assessing cognitive and language development in bilingual children.

Thus, this study offers a contribution to the state of the art of our understanding of the influence of living in bilingual/bicultural milieus on cognitive development because the six patterns found provide: (1) theoretical implications for overcoming methodological problems, and (2) practical implications for guiding further research on how to measure bilingual children's cognitive and language development for reaching accurate assessment and placement decisions.

Methodological Problems Affecting the Assessment of Bilingual Children

The methodological problems that will be critically reviewed in this article focus on the issue of controlling for external factors influencing

the valid and reliable assessment of bilingual low-income children. These important methodological problems have not yet been solved: the solutions often proposed, such as using translated versions of standardized tests, are inappropriate. It is critical to solve these methodological problems as currently validity and reliability are presumed to pertain to the educational, social, ethical, and moral consequences of using assessment instruments for accurately assessing and placing bilingual children (AERA, APA & NCME, 1985; Messick, 1989).

Given that state-of-the-art standardized instruments lack validity for bilingual children, a number of authors (e.g., Frasier, 1991; Gonzalez, 1991; Loyola, McBride, & Loyola, 1991; Oiler, 1991) have suggested ways to overcome present methodological problems: (1) to assess dual language proficiency; (2) to rely on non-verbal culturally appropriate tests rather than on verbal intelligence tests; (3) to use multiple measurements across cultural-linguistic contexts; (4) to improve the construct validity of assessment decisions by using individualized qualitative methods; and (5) to stimulate evaluators' awareness of the influence of their prior knowledge, conceptualizations of constructs measured, and cultural/linguistic backgrounds on assessment decisions of bilingual children. These suggestions given by major researchers, transformed into research questions, will be critically analyzed in light of literature below.

Does language of assessment influence bilingual children's performance? Gonzalez (1991,1994) constructed a model that explained cognitive and language development of bilingual children as a triple interaction between cognition, language, and culture. That is, she argued that living in bilingual and bicultural milieus influences cognitive development, and that cognitive development in turn influences first (L1) and second (L2) language proficiency. Gonzalez (1991, 1994,1995) demonstrated that using verbal and non-verbal classification tasks including stimuli groupings representing cultural semantic categories reflected in linguistic gender assignments gave bilingual children the opportunity to express different levels of conceptual abilities in relation to L1 and L2. Moreover, she found a difference in bilingual children's conceptual development in relation to the referent content category with children performing better on the verbal classification tasks for

inanimate (food) than for animate (animals) referents in both languages. As a result, we propose that the administration of QUEST in different languages, which reflects cultural semantic categories that may be similar or different across languages, allows evaluators to have a holistic view of the triple interaction of cognition, language, and culture.

This interaction between bilingual and bicultural milieus and cognitive development has been suggested by several authors. For instance, Cummins (1991) suggested that there are two aspects of language proficiency in bilinguals: (1) attribute-based aspects encompassing cognitive and personality variables, and (2) input-based aspects referring to the quality and quantity of input received from the environment. Cummins (1991) stated, "Moderately strong crosslingual relationships are observed for attribute-based aspects of L1 and L2 proficiency as a result of the fact that underlying attributes of the individual manifest themselves in the individual's performance in both languages" (p. 54). On the other hand, input-based academic and decontextualized aspects of L1 and L2 proficiency across a variety of sociolinguistic situations and ages show only consistent moderate relationships across languages due to the influence of cognitive individual characteristics of the learner. Moreover, according to Harley, Cummins, Swain, and Allen (1990), L1 and L2 proficiency include oral and academic aspects. They argued that bilingual children tend to show lower verbal than performance abilities when administered intelligence scales due to not yet having fully developed yet academic English language proficiency. These lower scores in verbal intelligence scales can be easily confounded with learning disabilities; consequently, bilingual children can be misplaced in special education, due to the fact that most of them also show oral (but not academic) English language proficiency as measured by language scales. These authors explained the overrepresentation of bilingual children in special education as the effect of a major problem with how dual language proficiency is conceptualized. That is, there is a failure to take into account an important factor explaining low scores on verbal intelligence scales: bilingual children's academic language proficiency in English.

In addition, several authors have also referred to practical implications of different conceptualizations of L1 and L2 proficiency on the accurate assessment of bilingual children. For instance, Ysseldyke

(cited in Reynolds, Wang, & Walberg, 1987) pointed out that if definitions of learning disabilities now in use in the U.S. were applied to limited English proficient students, more than 80% of them would be misdiagnosed. Hamayan and Damico (1991a), argued that bilingual children's oral communicative difficulties in English (e.g., reduced vocabulary, mispronunciations, grammar and syntax errors) may be confounded with genuine manifestations of language, speech, or learning disorders especially if cultural and linguistic differences are not taken into account. They suggested that genuine disorders can be accurately diagnosed when L1 and L2 assessment is conducted and problems are identified in both languages. As stated by Chamberlain and Medinos-Landurand (1991), "When assessing in only one language, a disability cannot be accurately distinguished from limited English proficiency" (p. 133); whereas "Comparing performance on tasks in both the L1 and L2 may well provide important clinical insights into a student's problem" (p. 134), especially for young bilingual students who come from low socioeconomic status households. Relatedly, Pollack (1980) found that when assessing bilingual students, the dominant language should be used for the first administration of tests in order to obtain optimum scores. Moreover, Commins (1986) found that when both languages were used simultaneously for assessing children with relatively low language proficiency levels in both languages, their performance improved in comparison to single-language administration. Furthermore, Chamberlain and Medinos-Landurand (1991) suggested that code switching be allowed when assessing bilingual children as they use rules of both languages and they should not be penalized if correct answers are provided in more than one language. If code-switching is not allowed, it can become "a barrier to the students' effectiveness in test-taking" (Chamberlain & Medinos-Landurand, 1991, p. 128). Is there a difference in bilingual children's performance when using verbal and non-verbal assessment procedures? The issue of how to measure verbal and non-verbal knowledge representation in bilinguals has become a traditional methodological problem. This difficulty is due to the presence of two verbal systems and two cultural symbolic systems that are hypothesized to interact in different ways with abstract knowledge representational systems. In order to illustrate the presence of different theoretical positions underlying our current assessment

problems of measuring cognitive and language development in bilingual children, we will briefly describe three approaches including the theory of semantics, the dual-code model, and variants derived from the dual-code model. A traditional approach to the study of the representation of knowledge is an early theory of semantics (Lambert, Havelka, & Crosby, 1958) with two traditional hypotheses: (a) the concept mediation hypothesis, with one non-verbal representational system for abstract concepts that are independent of both languages (leading to compound bilingualism); and (b) the word mediation hypothesis, with two verbal representational systems for words in both languages (leading to coordinate bilingualism).

Due to the inconsistent results obtained for the two traditional hypotheses, a second approach, the dual-code model (Paivio & Lambert, 1981) emerged. This model states that connections between concrete semantic representations in both languages should converge in common imaginal representations and be stronger than connections between abstract or affective semantic representations. As a third approach, several other variants of the dual-code model have appeared, some focusing on the links between verbal representational systems in both languages (see e.g., Kirsner, Smith, Lockhart, King, & Jam, 1984; Potter, So, Von Eckardt, & Feldman, 1984), and others focusing on the interconnection between the non-verbal and verbal representational systems in both languages (see e.g., Hakuta, 1976; Kellerman, 1983).

The differences hypothesized by these three approaches in how verbal and non-verbal knowledge is represented by bilinguals is also reflected in results of recent research studies. For instance, Chapman (1991) pointed out that studies measuring cognitive abilities through non-verbal methods have typically found children to perform at higher developmental levels than studies conducted within the traditional Piagetian approach that used predominantly verbal administration procedures. Chapman (1991) even went further, and stating that genuine cognitive developmental levels in children are traditionally underestimated by assessment procedures that rely only on verbal methods. He considered that verbal and non-verbal procedures are complementary, and not alternative, criteria for assessing children's cognitive development. This particular point was also made by Gonzalez (1991, 1994), who found that currently used standardized tests for

assessing language proficiency (e.g., IDEA, Ballard, Tighe, & Dalton, 1979), and even non-verbal intelligence scales (e.g., Test of Nonverbal Intelligence [TONI], Brown, Sherbenou, & Dollard, 1982) underestimated young bilingual children's language-cognitive abilities. Moreover, normal developmental levels were attained by bilingual children when assessed verbally, and superior developmental levels were attained when they were assessed non-verbally with QUEST. Thus, when using non-verbal qualitative assessments, bilingual children's higher conceptual developmental levels can be tapped. Furthermore, according to Gonzalez (1991, 1994, 1995), it appears that the process of verbal and non-verbal knowledge representation is not related to the language learning idiosyncratic experiences in bilinguals, leading to compound or coordinate bilingualism, as the two traditional hypotheses from the mediation theory of semantics proposed (Lambert, Havelka, & Crosby, 1958). Instead, as the data of Gonzalez's study (1991, 1994, 1995) showed, every bilingual child constructs verbal and non-verbal concepts at different developmental levels in relation to the particular cognitive, cultural, and linguistic characteristics of the content being learned. In relation to representational systems, Gonzalez's (1991, 1994, 1995) findings showed that verbal and non-verbal conceptual development is influenced by similar or different linguistic structures and markers, sociocultural meanings for linguistic conventions, and underlying abstract and semantic conceptualizations in the L1 and L2 and their cultures. Thus, both content and language of administration interact leading to different performance levels attained by bilingual children when measuring verbal and non-verbal knowledge representations.

Do multiple measurements and informants enhance construct validity in assessment decisions with bilingual children? A major and well-known assessment principle states that assessment, placement, and instructional decisions have to be based on a battery of measures including multiple informants and contexts (e.g., parents' and teachers' ratings, observations across different social and academic contexts, etc.). We argue that using qualitative assessments with bilingual children is very important because they tap the multidimensional interaction between cognition, language, and culture; in contrast, discrete point standardized tests measure language and cognitive development only

unidimensionally. Moreover, the methodological problems encountered when assessing bilingual children are related to the characteristics of traits or constructs measured. These abstract constructs are created by theorists within specific schools of thought, and they are complex omnibus concepts that cannot be measured thoroughly by discrete point standardized tests. In contrast, because qualitative assessment is based on a multidimensional developmental view, it provides verbal and non-verbal performance patterns, enhancing construct validity. These performance patterns can shed some light on the methodological problems of forming experimental groupings of bilingual children based on their language proficiency and dominance.

Moreover, several authors have highlighted the need to use multiple measurements across different contexts and informants for accurately assessing bilingual children's cognitive and language development. For instance, Harley et al. (1990) suggested that different measurements be used for determining L1 and L2 proficiency, because individuals have had unique experiences with differential effects on language proficiency. They argued that the language proficiency construct needs to take into account the developmental context of language learning. Relatedly, Lewis (1991) stated that comprehensive assessment has as a goal "to produce an accurate appraisal of students' current level and mode of intellectual functioning within the context of their cultural background and experience" (p. 127). Moreover, he pointed out that assessment is a continuous process that needs to be merged with instruction and that should be based on a wide range of informants including teachers, evaluators, parents, and school administrators. In addition, Samuda (1991) pointed out that no one standardized or qualitative measurement is sufficient to accurately assess bilingual children. He also referred to erroneous practices of evaluators that reach diagnostic decisions based on limited assessment data, and on lack of knowledge and misconceptions regarding assessment principles. Thus, both the use of a battery of measurements and different informants across cultural and linguistic contexts are important for enhancing construct validity when assessing bilingual children.

Does the use of individualized qualitative methods improve assessment decisions? We consider that it is very important to individualize assessment in order to tap idiosyncratic developmental

patterns in bilingual children. In relation to this argument, Cronbach (1986) pointed out that individuals interpret the same tasks or problem-solving activities in different idiosyncratic ways as portrayed by the different testing strategies. In addition, Sternberg (1991) pointed out that the same tasks administered to different individuals cannot tap their idiosyncratic abilities, and that particular tasks may influence the evaluator's interpretations and diagnostic decisions. Thus, idiosyncrasies present in the child and the evaluator influence the assessment process. We need to acknowledge the subjectivity involved in assessment by measuring idiosyncrasies present in bilingual children.

In addition, standardized tests do not tap idiosyncratic abilities in bilingual children, resulting in dangerous overgeneralizations of cultural patterns of minority groups that have become common misconceptions among educators and evaluators (see e.g., Gonzalez, 1993a, 1993b). For instance, the proliferation of the concept of learning styles results in dangerously oversimplified and overgeneralized stereotypical views of minority children, such as when all Hispanic children are portrayed as "context-dependent." This stereotypical view of Hispanic children prevents educators and evaluators from recognizing that, besides being raised in different linguistic and cultural milieus, minority children have idiosyncratic characteristics. We argue that bilingual children should be viewed as individuals, and not through stereotypical cliché's that deny their right to an equal treatment by evaluators and educators in assessment and instructional situations. Moreover, to give an equal education to minority children does not mean to treat every child in the same way, but to use appropriate assessment and instructional methods that match each child's individual abilities (Samuda, 1991).

Do evaluators' prior knowledge, conceptualization of constructs measured, and cultural/linguistic backgrounds influence assessment decisions in bilingual children? We have started to recognize that the evaluator's personality is a major assessment tool which may bring biases to the assessment process. That is, the evaluator's identity, values, beliefs, cultural stereotypes, life experiences within particular linguistic and cultural milieus, and prior knowledge on measured constructs influences assessment decisions (see Gonzalez, 1993a, 1993b; Hamayan & Damico, 1991b). Moreover, in the case of identifying gifted bilingual children, several authors (e.g., Frasier, 1991; Renzulli, 1991) have

pointed out the multidimensional definitions of giftedness reflected in current assessment instruments that are inappropriate for bilinguals. However, most educators and psychologists still believe that there must be a "panacea," a non-biased instrument that can be used by a trained evaluator to avoid cultural and linguistic biases in the assessment process (see Gonzalez & Yawkey, 1993). But, several authors (e.g., Roth, 1988) have pointed out the presence of a major flaw in this reasoning because an "objective" instrument cannot exist independently of the subjectivity brought to the assessment situation by evaluators. For instance, the presence of evaluators and other testing conditions can modify behaviors assessed.

In summary, it is our argument that qualitative assessments provide a holistic view that helps to understand and improve methodological problems still present when assessing bilingual children. We believe that using appropriate cultural contents and verbal and non-verbal problem-solving tasks administered in L1 and L2 gives bilingual children the opportunity to express their cultural and linguistic content knowledge associated with either of their languages. Moreover, we consider that the flexibility of individualized assessment adds a window for evaluators to observe idiosyncratic characteristics of bilingual children's thoughts. Thus, it is our argument that qualitative assessments provide evaluators with an insightful view of how the interaction of cognition, culture, and language is expressed in bilingual children's performance. Finally, including two independent evaluators for assessing the same child can capture the subjectivity involved in assessment, and can provide a holistic view of bilingual children as evaluators may highlight complementary observations and interpretations.

Research Questions

Five research questions guided the qualitative data analysis of this study: (1) Does language of assessment influence bilingual children's performance? (2) Is there a difference in bilingual children's performance when using verbal and non-verbal assessment procedures? (3) Do multiple measurements and informants enhance construct validity in assessment decisions? (4) Does the use of individualized qualitative methods improve assessment decisions? (5) Do evaluators' prior

knowledge, conceptualization of constructs measured, and cultural/linguistic backgrounds influence assessment decisions? In addition to an examination of these questions, the interaction across patterns found will be explored.

Method

Subjects

Subjects for this study were 17 Hispanic bilingual 5-6 year-olds, 1st, 2nd, and 3rd generation Mexican-Americans, attending a public school located in a low socioeconomic status neighborhood. Subjects were selected based on two referral procedures for giftedness including: (1) the Home Language Survey (HLS, Gonzalez, 1991), and (2) a locally-designed Teachers' and Parents' Rating Scale of Creativity (TPRSC). Eleven boys and 6 girls were selected who were attending a half-day monolingual English (8 children) and bilingual Spanish/English (9 children) kindergarten in a metropolitan Southwest school district with 50% minority students. Children were considered unbalanced bilinguals when they used both languages at home and school, even if they were not proficient in either language.

Instruments

Home Language Survey. The HLS was developed by Gonzalez (1991) and consists of open-ended questions and Likert scales asking parents to: (1) self-rate their Spanish and English proficiency, (2) rate their child's Spanish and English proficiency, (3) rate the frequency of use at home of Spanish and English, (4) provide the ages of the children in the home, and (5) provide background information of both parents (i.e., their birthplace, how many years they have lived in the U. S., and their occupations and educational levels).

Teacher's and Parents' Rating Scale of Creativity. The TPRSC was locally-designed (Fleming, Gonzalez, Maker, Nielsen, & Rogers, 1992) and consisted of seven open-ended questions asking respondents to describe the child's linguistic, problem-solving, and individual and group working and playing abilities at home and at school. Finally, parents and teachers were asked to circle descriptors for the child's abilities, and write additional comments.

Qualitative Use of English and Spanish Tasks (QUEST). Gonzalez's (1991, 1994, 1995) developed a model that explained the concept formation process in bilingual children and identified two knowledge representational systems dependent on the particular cognitive, linguistic, and cultural characteristics of the content learned. The first conceptual representational system is abstract, universal, and non-verbal; the second is semantic, verbal, and culturally-linguistically bound. Cognitive factors were considered abstract knowledge representations instantiated in cultural symbolic conventions and in linguistic structures and markers. Cultural and linguistic factors were selected because Spanish assigns linguistic gender for both animate and inanimate abstract conceptual categories, corresponding to culturally important symbolic distinctions, that are expressed through linguistic rules and markers. In contrast, English only assigns linguistic gender to some animate conceptual abstract categories. The model from which the classification tasks were derived was based partially on Piagetian theory (Piaget, 1967) and on the constraint model (Markman, 1984; Waxman, 1990), and it was found to have construct validity as shown by parametric and non-parametric tests (Gonzalez, 1991, 1994, 1995). The derived verbal and non-verbal classification tasks were designed to assess bilingual children's general and linguistic gender conceptual processes for two different abstract, symbolic, and linguistic semantic categories represented by animals (animate) and food (inanimate) objects. Stimuli used for the five classification tasks were plastic full-color objects representing 14 groupings reflecting the interaction of cognitive, cultural, and linguistic factors. Stimuli groupings were validated using judges for assuring construct validity and three pilot tests for assuring content validity (Gonzalez, 1991, 1994, 1995).

Moreover, internal validity and reliability of the classification tasks were demonstrated by using Pearson chi-square association tests in this study in order to control for the effect of using two sets of stimuli. Two sets of two-way tables were tested, one for Spanish and one for English, using the five verbal and non-verbal classification tasks divided by the two referents (animals and food) and the two scoring areas of point assignments (general and gender) versus the two stimuli sets. All but three of the two-way tables were not significant, indicating that using two different sets of stimuli was not associated with the developmental

level at which bilingual children performed on verbal and non-verbal classification tasks. The three tables that were significant included the following task administered in English: (1) the labeling task for food referents and the general area of point assignments ($p < .020$), (2) the labeling task for animal referents and the gender-based area of point assignments ($p < .047$), and (3) the category clue sorting task for animal referents and the gender-based area of point assignments ($p < .008$). These three tables indicate that for the English administration set 2 was easier than set 1, because children attained higher developmental levels for set 2 than for set 1. We interpret these results as the effect of level of familiarity of the children with the stimuli included for the labeling and category clue sorting tasks in English for sets 1 and 2.

Three of these five classification tasks are verbal including labeling, defining, and verbal justification of sorting; and two tasks are non-verbal including sorting and category clue. Tasks will be described following the pre-established order of administration (for a more complete description of tasks see Gonzalez, 1991, 1994, 1995). For the labeling task, the child is presented plastic objects and asked to name them (*What do you call this?*), while being given one item at a time, followed by the defining task at the production level in which the child is asked four different probes to elicit a description of the object(s) (*What is a __? What is a __ like? Tell me something about a __. What does a __ look like?*). To tap the comprehension level of the defining task, the child is then given a definition that points to verbal and non-verbal clues for class inclusion categories of objects (taxonomic categories: superordinate, intermediate, and subcategories). This definition is repeated three times, after which the child is asked to define three different kinds of items. For the sorting task, the child is asked to group the objects by linguistic gender; followed by the verbal justification of sorting task in which the child is asked to explain the order imposed on the objects, and she is presented with metalinguistic counterexamples that change groupings and labels. Finally, for the category clue task, the child is provided with a model of how to group objects by linguistic gender using two pictures of identical dolls, and then is asked to sort the objects following the model provided, to explain the groupings, and to answer metalinguistic counterexamples that change groupings and labels.

The scoring system is divided into five point assignment areas including language development, verbal and non-verbal general, and verbal and non-verbal gender areas, based on which children are diagnosed on conceptual development (for an extended description of scoring see Gonzalez, 1991, 1994, 1995). General areas include any valid criteria that the child uses for classification (e.g., color, functions, subcategories, etc.). Gender areas include classification criteria based on physical gender for animates, linguistic gender assignment for inanimates, or functional use for both animates and inanimates. The language development area includes only the labeling task. The verbal general and gender-based areas include defining and verbal justification of sorting tasks, and the non-verbal general and gender-based areas include sorting and category clue tasks. Thus, children's responses to the five tasks administered in both languages were scored twice, assigning points for both general and gender areas. The language development area was categorized into three levels: (a) low (0-2 points), (b) moderate (3-5 points), and (c) high (6-8 points), according to the number of labels produced by the child. For the other four areas children's responses were categorized into five stages based partially on Piaget's theory (1965): (1) no classification (affective responses, juxtaposed groupings and graphic collections), (2) pre-conceptual: perceptual (extralinguistic features - color, size, shape, parts of objects), (3) pre-conceptual: functional (thematic relations), (4) concrete (taxonomic categories showing class-inclusion), and (5) metalinguistic (taxonomic semantic categories). In order to be diagnosed in any of these five developmental stages for any of the five verbal and non-verbal tasks, children's responses needed to be at that level at least for three out of the eight items that were included in the tasks.

Procedure

A large metropolitan school district in the Southwest with 50% minority students adopted QUEST as an alternative individualized procedure for selecting and placing bilingual Hispanic students in gifted education. Two parallel sets of stimuli presented in a consistent sequential order were used to avoid transfer of learning effect. In addition, Spanish and English language administrations of approximately an hour each were used within a three-week interval for

the majority of children (11), and for the other six children it was administered within a two-month interval. Eleven children were administered QUEST in both languages, allowing children to respond in their dominant language; this resulted in three patterns: (1) all eight children assessed first in English and then in Spanish responded in English to the English administration, and two responded in English to the Spanish administration; and (2) all three children who were assessed first in Spanish and then in English responded in Spanish to the Spanish administration, and one responded in Spanish to the English administration. In addition, six children were evaluated twice in the same language (one in Spanish and five in English) because children could not understand the questions asked in their less dominant language.

Fourteen graduate students majoring in educational psychology who were completing a course on testing of minorities served as evaluators. They received ten hours of training in administering QUEST, including watching three administration videos, and hands-on practice sessions where pairs of students administered the assessment to each other and scored protocols. Two pairs of evaluators were trained in Spanish, each group consisting of a native Spanish speaker who administered QUEST, and a non-native Spanish speaker who served as the recorder. The other five pairs of evaluators were trained in English, each pair consisting of two monolingual English speakers who served as administrators and recorders. Each pair assessed two children, recorded responses in protocols, made an assessment decision on the children's verbal and non-verbal conceptual developmental levels, and wrote a summary report. The second and third co-authors of this paper formed a Spanish pair of evaluators and integrated the summary reports and protocols of the two independent evaluations conducted for each child into psychoeducational reports including assessment and placement decisions. QUEST reports and the referral data were examined by an interdisciplinary placement committee formed by teachers, administrators, school psychologists, parents, university faculty, and graduate students.

Description of Data Analysis

Qualitative analysis of data included the complex scoring system for QUEST, case studies, and Pearson chi-square tests of association. When using the scoring system, children's responses were assigned nominal categories that described their verbal and non-verbal conceptual development achieved across classification tasks. These nominal categories were used for the insightful interpretation of case studies and chi-square tests. Single and multiple case studies were used as a learning method for understanding the complex influence of language and culture on cognitive development, for which extensive description, and holistic and contextual analyses were conducted (US General Accounting Office, 1987). We used 17 single-case studies for analyzing the complex interaction of multiple qualitative measures and informants, forming a unique group well-suited for this design (Yin, 1994). In addition, multiple case-studies were used for addressing validity issues: (1) selecting a heterogeneous sample of 17 bilingual children with different Spanish and English proficiency levels for increasing the probability of finding implications not predicted by the original model (Yin, 1994); (2) examining children's responses recorded in the QUEST protocols for identifying emerging similar patterns across cases using "the chain-of-evidence technique in data reduction" (US General Accounting Office, 1987, p. 38); (3) using the triangulation method (Stake, 1988; Yin, 1994) for examining findings in QUEST documented by two or more evaluators, and parents' and teachers' surveys. The three co-authors of this paper independently found similar patterns across the 17 cases from their own observations, the protocols, and written summary reports. Then, the emergent patterns were developed into theses for this study.

For the Pearson chi-square tests of association, a conservative .01 alpha level for the overall test of significance was set due to the number of tests performed. Since the single-case study analysis showed different developmental levels achieved by children for the non-verbal and verbal components of the category clue task, they were treated separately in the analyses. Thus, while five tasks were used when constructing the Gonzalez model (1991), this study categorized the data into six tasks by splitting the category clue task in two components: sorting and verbal justification. Moreover, because not all students were assessed in both languages, resulting in unequal numbers for each language, chi-squares

were performed separately for each language of administration. When the administration of QUEST was conducted twice in the same language, the first administration was used for statistical analysis. However, when the first assessment had missing data, the second complete assessment was used. These data reflect the reality faced by evaluators when assessing bilingual children who are not fully proficient in both languages, and who do not have either language dominant.

Thus, sixteen two-way tables clustered in two sets of chi-square tests of association were conducted, in which language of administration and animals and food referents were kept separate, including: (1) conceptual development achieved for general and gender-based areas of point assignments versus conceptual development achieved for sorting, verbal justification of sorting, category clue sorting, and category clue verbal justification tasks; (2) conceptual development achieved for non-verbal tasks (sorting and category clue sorting) versus conceptual development achieved for verbal tasks (verbal justification and category clue verbal justification). In addition, 8 two-way tables clustered in a third set of chi-square tests of association were conducted, in which language of administration and general and gender-based point assignments areas were kept separate for analyzing, including (3) conceptual development achieved for animal and food referents versus conceptual development achieved for labeling, definition, sorting, verbal justification of sorting, category clue sorting, and category clue verbal justification tasks. A summary of the chi-square tests of association, degrees of freedom, Cramér Vc statistic, and level of significance for the 24 two-way tables for Spanish and English, animal and food referents, and general and gender-based point assignment areas is presented in Table 3. Due to space limitation, only significant tests are included. Limitations of data analysis relate to the number of chi-square tests run and the number of subjects included.

Results and Discussion

First, we will provide a summary of descriptive data regarding the conceptual developmental stages attained by the 17 bilingual Hispanic kindergartners assessed for the five verbal and non-verbal classification tasks of QUEST across the two sets of stimuli, animals and food, and the

two languages of administration, Spanish and English. These descriptive data can be useful for identifying the conceptual developmental levels at which the referred children were performing. Second, we will present the six patterns found in multiple-case study analyses that have theoretical and practical implications for the assessment of bilingual children's cognitive and language development. In addition, excerpts from single-case studies will be used for illustrating how all six patterns emerged from rich qualitative measures. The first two patterns will be supported also by significant chi-square test results. In order to provide consistency across patterns, one case of Verónica, a third generation Mexican-American six year old girl, will be presented. Pseudonyms are used for all case studies to protect children's identity. For a summary version of the theoretical and practical implications of these six theses see Gonzalez, Bauerle, and Felix-Holt (1994b).

Description of Conceptual Developmental Stages Attained by Bilingual Children

Providing descriptive data of the conceptual developmental stages attained by the 17 bilingual Hispanic kindergartners assessed for the five verbal and non-verbal classification tasks of QUEST can be useful for identifying children performing at average, below average, and superior/gifted developmental levels. As the summary presented in Table 1 (see Appendix) for verbal tasks shows, children performed better in the labeling task in Spanish for both animals and food referents, and also for both general and gender-based areas of point assignments. For this labeling task most children performed at the pre-operational and concrete levels across referents and languages of administration, and the metalinguistic level was achieved only by some children in Spanish for food referents. For the defining task for both general and gender-based areas of point assignments, children performed better in Spanish for animal referents. Most children performed at the pre-operational level for the defining gender task, and at the pre-operational and concrete levels for the defining general task. The metalinguistic level was achieved only by some children for the definition general task. For the verbal justification of sorting task for both general and gender-based areas of point assignments, children performed better in Spanish for the food referents, followed by Spanish and English administrations for

animal referents. For this task, most children performed at the pre-operational and concrete levels, and the metalinguistic level was achieved only by some children in the Spanish administration for food referents.

The summary provided in Table 2 (see Appendix) for non-verbal tasks shows that for the category clue verbal justification of sorting task for both general and gender-based areas of point assignments, children performed better in Spanish for food referents. Moreover, most children performed at the pre-operational level in this task. For the sorting general and gender based areas of point assignments, children performed higher in the Spanish administration for food referents. Most children performed at the concrete level for the sorting general task, and only four children reached the metalinguistic level in Spanish for food referents. In contrast, most children performed at the pre-operational level for the sorting gender task. For the category clue sorting general and gender areas of point assignments, children also performed higher in Spanish for food referents. Moreover, for the category clue sorting general task most children performed at the concrete level, with few reaching the metalinguistic level. For the category clue sorting gender task, most children performed at the pre-operational level with few achieving at the metalinguistic level for the Spanish administration for food referents.

Thus, in summary across verbal and non-verbal tasks children performed at higher developmental levels in the Spanish administration for food referents, even reaching the metalinguistic level; and children also performed at higher levels for general in comparison to gender-based areas of point assignments. Moreover, the most common levels attained by children was pre-operational for gender-based areas of point assignments, and concrete for general areas of point assignments for both verbal and non-verbal tasks. In relation to the chronological age of these children (approximately an average of 5 years and 6 months), the pre-operational level achieved most commonly is considered developmentally normal (Piaget, 1967), and the concrete level is considered developmentally superior for kindergartners. Finally, there was a tendency in children, shown consistently across tasks, to be able to achieve at the metalinguistic level in the Spanish administration with food referents. Thus, the domain of cultural knowledge tested and the

language of administration of tasks were influencing the developmental level attained by these bilingual children, as shown by the two areas of point assignments.

Language of assessment made a difference in bilingual children's performance

Multiple-case study analyses and significant results found in the first set of chi-square tests (see Table 3) support our first pattern that language of administration made a difference in cognitive developmental levels achieved by bilingual children. The first set of chi-square tests was performed with the purpose of analyzing whether there was an association between children's conceptual development achieved on gender-based linguistic concepts and on general concepts across the six tasks administered in English and Spanish using animal and food referents. For the English administration of QUEST, category clue sorting for food referents ($p < .01$) and category clue verbal justification for animal referents ($p < .5$), were found to be statistically significant when general and gender-based point assignments were compared. When QUEST was administered in Spanish, category clue sorting ($p < .5$) and category clue verbal justification ($p < .5$), both for animal referents, were found to be statistically significant when general and gender-based point assignments were compared. Category clue sorting for food referents, when QUEST was administered in English, made a difference as bilingual children could not attain the highest metalinguistic conceptual level in which linguistic and semantic categorizations are used because English does not assign linguistic gender for food referents present in Spanish. Category clue sorting administered in Spanish, and category clue verbal justification administered in English and Spanish, all for animal referents, made a difference because bilingual children could sort and verbally explain their conceptual groupings in a similar manner in both languages by using natural physical gender for general classifications.

Based on the multiple-case study analysis, the first thesis of this first patterns refers to higher cognitive developmental levels attained by bilingual children, even by balanced bilinguals, when QUEST was administered in Spanish. These findings can be explained as the result of Spanish providing bilingual children the opportunity to express different

cultural and linguistic conceptual representations, not present in English, that enhanced their performance. Therefore, adding an L2 administration opens new, culturally and linguistically bound conceptual dimensions that could not be observed in the L1 administration. For instance, three balanced bilingual children who were assessed with QUEST in Spanish performed at higher cognitive developmental levels than when they were assessed in English. During the Spanish administration these three balanced bilinguals attained the highest metalinguistic conceptual developmental level as they sorted food objects according to linguistic gender. In contrast, their performance in English when sorting food objects attained only the concrete conceptual developmental level. Thus, assessing bilingual children using two languages of administration is not repetitive but complementary, because culturally and linguistically appropriate qualitative methods evoke the positive influence of two cultures and languages on cognitive development.

The second thesis of this first pattern highlights the importance of allowing non-balanced bilingual children to use code switching when assessed with QUEST for enhancing their performance on cognitive and language development. This practice has been supported by several researchers (e.g., Chamberlain & Medinos-Landurand, 1991) for avoiding barriers when assessing bilingual children. In this study, when assessed in English, four children responded in Spanish; and when they were assessed in Spanish, one child responded in English. This shows that almost one third of the children assessed (5 out of 17 cases) could perform the required tasks even when assessed in their non-dominant language, as long as they were allowed to respond in their dominant language. For example, Verónica (an English-dominant bilingual child) responded in English to both Spanish and English administrations. When Verónica was assessed in Spanish, she achieved the metalinguistic level in the verbal justification task for animal and food referents, and also in the category clue task for food referents. However, Verónica performed at the concrete level on all tasks when she was assessed in English. This case demonstrates that Verónica's potential to perform at the highest metalinguistic level was tapped only by the Spanish administration for which she was allowed to respond in English and to use code switching, highest level that could not be tapped when she was assessed in English. This second pattern points out that dual language assessment is not only

important for bilinguals who show low language proficiency levels in both languages (as suggested by Cummins, 1991), but also for bilinguals who show a clear language dominance (as illustrated by Verónica's case).

Several authors (e.g., Chamberlain & Medinos-Landurand, 1991; Cummins, 1991; Hamayan & Damico, 1991a; Harley et al., 1990; Pollack, 1980) have pointed out that the multidimensionality of the constructs of language proficiency and dominance cannot be measured accurately when using discrete point standardized tests that only measure oral, but not academic, language proficiency. We propose that academic language proficiency is related to semantic conceptual development in both languages. Therefore, measuring this construct with qualitative assessment methods in both L1 and L2, such as QUEST, provides important information for differentiating between normal L2 learning and genuine handicapping conditions and disabilities. Then, dual language administration of QUEST can more closely measure the multidimensionality of the triple interaction of language, cognition, and culture, and can shed some light on how cognitive development is influenced by bilingual/bicultural milieus.

Using verbal and non-verbal procedures provided new and valuable information

Significant results obtained in chi-square tests and multiple-case studies support our second pattern that using verbal and non-verbal classification tasks allowed evaluators to gather complementary information. The first thesis of this second pattern refers to the addition of a different assessment dimension by tapping both verbal and non-verbal conceptual processes. That is, it is our argument that bilingual children learn two languages that do not always coincide in how concepts are represented verbally by linguistic structures and markers, and non-verbally by cultural symbolic meanings and linguistic conventions, and abstract semantic categories. We found that assessing bilingual children using verbal and non-verbal gender categories made a difference in their performance in classification tasks. The second set of chi-square tests was performed to test whether there was an association between children's performance on non-verbal and verbal tasks administered in English and Spanish using animal and food referents for

general and gender-based point assignment areas (see Table 3 in Appendix). For this second set of tests, the labeling and defining tasks were excluded since they had no parallel non-verbal counterparts. For the English administration, a significant difference was found in children's conceptual development for sorting versus verbal justification of sorting for animal ($p < .05$) and food ($p < .05$) referents and general point assignment areas, and for animal ($p < .01$) and food ($p < .001$) referents and gender-based point assignment areas. Moreover, for the English administration, a significant difference was found for category clue sorting versus verbal justification of sorting for animal ($p < .5$) and food ($p < .5$) referents and general point assignment areas, and for animal ($p < .01$) and food ($p < .001$) referents for gender-based point assignment areas. When administered in Spanish, a significant difference was found in children's conceptual development for sorting versus verbal justification of sorting for animal ($p < .05$) and food ($p < .01$) referents and gender-based point assignment areas. Finally, when administered in Spanish a significant difference was also found for category clue sorting versus verbal justification of sorting for food referents, and general ($p < .5$) and gender-based ($p < .01$) point assignment areas.

Thus, bilingual children attained different cognitive developmental levels in verbal and non-verbal tasks for both general and linguistic-gender categorizations because in English linguistic gender is assigned only to some animate referents. That is, when children's English responses were scored twice, most of the time gender-based categorizations were not present. Thus, children were only given credit for general categorizations. In contrast, in Spanish linguistic-gender is assigned to almost all animate and inanimate referents, which was related to different verbal and non-verbal cognitive developmental levels attained by bilingual children in producing naturally gender-based categorizations as elicited by the sorting task. However, when assessed with the category clue sorting task in Spanish that provided categorization models for the child to follow, bilingual children's verbal and non-verbal conceptual development differed for both general and gender-based categorizations. This difference in bilingual children's performance in Spanish may be due to their ease in using gender-based categorizations in productive non-verbal tasks, such as sorting, and their difficulty in following the linguistic-gender assignment model provided

in comprehensive non-verbal tasks, such as category clue sorting, resulting in general categorizations. Thus, we argue that having knowledge of linguistic gender assignments in Spanish was a necessary but not a sufficient condition for performing at high levels in the non-verbal classification tasks. That is, some bilingual children could understand linguistic gender assignments, achieving at higher developmental levels non-verbally than verbally. However, other bilingual children could not understand linguistic gender assignments non-verbally even when knowing the linguistic markers for gender in Spanish.

For example, Gregg performed at higher developmental levels for non-verbal tasks than he did for verbal tasks in both Spanish and English. In addition, his non-verbal concepts seemed to be quite advanced for his age as when classifying animals he created parallel lines for gender with subgroups representing parents and babies. Moreover, three other children, Linda, Silvia, and Oscar, were able to sort the food according to linguistic gender, but they were unable to justify their classifications. Similarly, a monolingual Spanish-speaking child, Samuel, was able to sort the animals by habitat, demonstrating a concrete non-verbal concept formation ability, but he was unable to justify verbally his classifications in Spanish, attaining only the functional level.

In general, findings suggest that bilingual children could sort objects using general criteria in English and gender-based criteria in Spanish, but they could not verbally explain their groupings with the same easiness in both languages. Moreover, findings show that balanced bilinguals performed on non-verbal tasks at higher conceptual developmental levels than in verbal tasks in both languages. However, children who were clearly dominant in one language attained equivalent or higher conceptual developmental levels for verbal than for non-verbal tasks. For instance, Verónica's (an English-dominant bilingual child) showed a higher conceptual developmental ability verbally than non-verbally. While she could sort food into concrete categories, showing an above-normal non-verbal concept formation ability, she was able to provide verbal justifications at the abstract metalinguistic level, showing a superior verbal concept formation ability.

A second related thesis found in this second pattern was that the content used for assessment also made a difference in bilingual children's performance showing that when they were familiar with the assessment stimuli, such as food, they tended to perform at higher non-verbal conceptual levels. The third set of chi-square tests supports this pattern as it examines the association between children's conceptual development achieved across verbal and non-verbal classification tasks when different cultural contents (i.e., animal and food referents) representing symbolic and linguistic differences between Spanish and English were used. For the English administration, labeling ($p < .001$) and defining ($p < .5$), both for gender-based point assignment areas, were found to be statistically significant when different cultural contents were compared. There were no differences for the Spanish administration. Thus, the third set of chi-square tests shows that the presence or absence in English of linguistic gender assignments for animate (i.e., animals) and inanimate (i.e., food) referents influenced the conceptual developmental level attained by bilingual children. Results were not significant for the Spanish language as it assigns linguistic gender for both animates and inanimates.

Thus, when both the second and third theses of this second pattern interacted, the Spanish administration in relation to non-verbal tasks for food referents created for bilingual children an opportunity to think at a higher metalinguistic level. The Spanish language added three different dimensions, in comparison to the English language, stemming from abstract categories reflected in linguistic gender assignments and cultural symbolic conventions. This interaction of the second and third theses coincides with previous results obtained (Gonzalez, 1991, 1994, 1995; Gonzalez et al., 1994a) which showed that bilingual children performing at above-normal developmental levels non-verbally and at normal developmental levels verbally in relation to the specific cognitive, cultural, and linguistic characteristics of the content being learned. In relation to the traditional problem of knowledge representation in bilinguals, findings in this study provide additional support for Gonzalez's model (1991, 1994, 1995) revealing two interconnected representational systems, one for universal non-verbal concepts, and a second one for culturally-linguistically bound semantic concepts. Thus, it can be concluded that using a qualitative assessment

method that reflects content that may be similar or different across languages allows evaluators to gain a more holistic view of the interaction of cognition, culture, and language.

Multiple measurements and informants enhanced construct validity

Theses found in multiple-case studies support the third pattern referring to the need for evaluators to use a multidimensional battery and multiple informants for assuring the accurate assessment of bilingual children. The first thesis of this third pattern states the importance of using multiple qualitative methods and informants for measuring language proficiency and dominance, and concept formation in bilingual children. A traditional methodological problem of research studies with bilingual children has been finding a valid and reliable instrument for grouping them based on their language proficiency and dominance. Using discrete point language proficiency standardized tests has not solved the problem. We found that using bilingual children's L1 and L2 and verbal and non-verbal performances in QUEST, in relation to parents' and teachers' ratings of children's language proficiency at home and at school, can help to accurately assess their language proficiency and dominance.

The second thesis of this third pattern points to the value of using multiple measurements and informants within qualitative assessment paradigms to gain a more holistic and dynamic view of how culture and language interact with cognitive development in bilingual children. Thus, as pointed out by several researchers (e.g., Harley et al., 1990; Lewis, 1991; Samuda, 1991), employing multiple measurements and informants leads to enhanced concurrent and construct validity of assessment instruments in bilingual children. Concurrent validity is improved by tapping different and complementary dimensions of the interaction of cognitive and linguistic abilities with different cultural content dimensions. At the same time, construct validity is enhanced by conceptualizing cognitive development as a dynamic learning process influenced by language development within a bicultural environment. That is, we consider that children's potential for learning is nurtured by living within a multicultural and bilingual milieu.

Verónica's case is used to illustrate the two theses of this third pattern. Her parents completed the HLS and rated her English

proficiency as "average for her age," and her Spanish proficiency as "not quite average for her age." The language reported by Verónica's parents as being most often used at home and among her peers was English, but some Spanish was spoken at home with Verónica's grandmother. Thus, as reported by her parents, Verónica's dominant language was English, so she was administered QUEST in English, and then in Spanish with a parallel set of stimuli by two different bilingual evaluators. Two independent pairs of evaluators administered QUEST to Verónica. Even though they were using different languages, they reported similar findings, providing support for concurrent validity and reliability. When taking into consideration both independent evaluations, Verónica demonstrated verbal concept formation abilities that surpassed those of her peers. Moreover, results from the parent and teacher surveys provided additional construct validity for the results obtained in QUEST allowing a more holistic view of the child. The parent survey demonstrated Verónica's talents and interests in verbal, cognitive (i.e., visual, auditory and spatial memory, spatio-temporal relationships), and social areas. The teacher's survey showed Verónica's strengths in mathematics, language, and social skills. She was described by both her parents and teacher as an inquisitive and independent worker, and a leader. However, contrary to the high ratings received by her parents and teacher, Verónica attained only a 61 percentile rank on the Raven's Coloured Progressive Matrices (Raven, 1976), which was far below the 97 percentile needed to enter the gifted program. Even though Verónica performed low on this standardized measure of non-verbal intelligence, when assessed with QUEST she showed normal non-verbal and above-normal verbal concept formation abilities for her chronological age in relation to Piagetian theory (1967). Findings in this case are typical of the bilingual children who participated in this study and support previous studies conducted by Gonzalez (1991, 1994). Thus, using multiple measurements and informants enhanced concurrent and construct validity, providing reliability of results and contributing to a more accurate assessment of Verónica's abilities.

Using individualized qualitative methods improved assessment decisions

Multiple-case study analysis supports the fourth pattern: there is need to individualize assessment for measuring unique individual,

cultural, and linguistic characteristics present in bilingual children. The first thesis of this fourth pattern refers to the value of qualitative methods for individualizing assessment as has been pointed out by several authors (e.g., Cronbach, 1986; Sternberg, 1991). In the case of bilingual children qualitative assessment provides a window to observe the presence of different content knowledge domains learned, and of cognitive processes that are nurtured by bilingual and bicultural milieus. For example, Verónica was allowed to respond in English when QUEST was administered in Spanish. Thus, as pointed out by several authors (e.g., Cronbach, 1986; Samuda, 1991), by adapting administration procedures to suit children's individual needs and abilities, evaluators allow children to demonstrate their understanding of different languages and cultures.

When using qualitative assessment procedures that take into account background experiences of bilingual children, cognitive and linguistic abilities that have been developed and learned within a specific cultural and linguistic environment can be observed. For example, one monolingual Spanish-speaking child, Samuel, when asked to group the animals by gender, grouped the animals by habitat. Although Samuel did not follow the directions for this task, he clearly demonstrated concrete thinking and extensive background knowledge, which was further demonstrated in his performance on the defining task. When Samuel was asked to define different animals, he described in detail with almost complete accuracy where they lived and how they cared for their young. This same point of above-normal concept formation skills in relation to chronological age is also well illustrated by Irene's performance, who responded in Spanish to the English administration. For the animal sorting task, Irene first created two vertical lines classifying animals by gender, and then immediately regrouped the animals into two piles: "los malos" (the bad ones) and "los buenos" (the good ones), showing flexibility of thinking.

Evaluators' prior knowledge, conceptualization of constructs measured, and cultural/linguistic backgrounds influenced assessment

Multiple-case study analyses were used with the purpose of supporting the fifth pattern: assessment is influenced by the subjectivity of evaluators' personalities. The first thesis of this fifth pattern refers to

our argument that the assessment of bilingual children is influenced by evaluators' idiosyncratic characteristics such as their cultural-linguistic background and prior knowledge, and conceptualizations of constructs measured. As a result, assessment becomes a subjective process, even in the case of administering standardized tests. In order to support this fifth pattern we are going to discuss experiences of the second and third Co-authors of this paper, who administered QUEST. When Daniel was assessed independently by these two evaluators in English, there were some similar observations including his short verbal and non-verbal responses, his lack of manipulating animal and food objects, and his animal groupings using the same number of objects present in the examples rather than by kind. However, there were also differences in how evaluators perceived Daniel's performance. One evaluator perceived Daniel as creative and efficient and seemed to be influenced largely by his behaviors showing social intelligence. The other evaluator perceived Daniel as quiet and serious as she reported that even though David was cooperative, he was not willing to elaborate on his responses.

These two evaluators also jointly assessed Verónica in Spanish and were able to interpret her behaviors differently because of their different linguistic backgrounds and conceptualization of metalinguistic ability (Chapman, 1991). When Verónica was asked a set of metalinguistic questions, she consistently responded, when probed, that the linguistic gender of the Spanish words could be changed. The evaluator who had acquired two languages simultaneously during childhood interpreted Verónica's responses as not demonstrating metalinguistic awareness since the child was not knowledgeable about social conventions for assigning different linguistic genders to objects. In contrast, the evaluator who had acquired Spanish as a foreign language interpreted Verónica's responses to probes as possible evidence of metalinguistic awareness. In this case, the child's lack of concern for linguistic gender changes was interpreted as an expression of her awareness of the arbitrariness of linguistic gender and of the independence of the meaning from the object's label. Thus, the presence of two independent evaluators is beneficial for assessing a child's performance because new observations and different perceptions and interpretations can be added. Furthermore, no evaluation should be interpreted without data from other informants.

Interaction across patterns and theses found

Multiple and single case-study analyses supported interactions across the six patterns and derived theses found. Results showed that QUEST provides bilingual children with the opportunity to perform differently when being assessed using two languages and cultural contents. Moreover, the multidimensionality of QUEST allowed evaluators to individualize assessment by providing a window for observing the interaction of cognition, language, and culture. Furthermore, the presence of two evaluators contributed with complementary assessment and placement "holistic" decisions for bilingual children as their observations were interpreted differently. Thus, we have argued across the six patterns and derived theses that qualitative assessment provides a holistic perspective that helps to understand and improve traditional methodological problems still present when assessing bilingual children.

The interaction of the different patterns and theses discussed above will be illustrated using information from the three case studies. Verónica's conceptual strengths seemed to be best tapped when QUEST was administered in Spanish as long as she was allowed to answer in English. By adapting the administration of QUEST to Verónica's verbal and non-verbal abilities in relation to abstract metalinguistic concepts, and by using other multiple qualitative measures and informants, we showed that in comparison to her bilingual peers, she demonstrated strength in verbal conceptualization. Finally, using evaluators with different backgrounds and prior knowledge enriched the interpretation of Verónica's verbal and non-verbal responses in the two languages of administration. For the other two case studies, Linda and Silvia, their performance were highest in the non-verbal tasks for food referents when administered in Spanish. Even though Linda responded mostly in English, the Spanish administration gave her the opportunity to demonstrate her accurate knowledge of cultural-linguistic gender concepts in Spanish, and the English administration gave her the chance to verbally justify her non-verbal responses at higher levels. The combination of these two administrations showed that both verbal and non-verbal conceptual development are the strengths of Linda, and that her opportunity to demonstrate her abilities depended upon the language of administration. Thus, individualizing assessment allowed Linda to

respond in either language, even when culturally and linguistically bound concepts were being assessed. In addition, when QUEST was administered in Spanish, Silvia demonstrated her accurate knowledge of linguistic gender assignments based on abstract concepts that are linguistically and culturally bound. When requested to sort food objects by linguistic gender assignments, Silvia was able to do so non-verbally and also to verbally justify her groupings by explaining that objects in each group were the same. Moreover, Silvia was able to transfer this culturally and linguistically bound concept of linguistic gender assignments when administered another non-verbal sorting task (i.e., category clue) with different food objects during the English administration. Thus, using this example we can observe that this second evaluation of Silvia adds insight into how the interaction of cognition, language, and culture is expressed in her cognitive performance. Furthermore, Silvia's performance was facilitated by the method's flexibility in allowing encouragement to be provided to her whenever she answered "I don't know," which eventually resulted in Silvia's production of verbal responses that were quite advanced for her age.

In addition, using two independent evaluators demonstrated the reliability of QUEST in producing complementary observations of each child's performance. For Linda, the first evaluator highlighted her cognitive flexibility, as she provided verbal justification whenever the evaluator changed and probed her groupings. The second evaluator, however, emphasized that Linda demonstrated consistency when verbally justifying her non-verbal categories. Therefore, the results of both evaluators' observations culminated in a more "holistic" view of the child. For Silvia, despite her limited verbal responses in both languages, she clearly transferred concepts from one language to the other and performed better when she became familiar with the evaluators, tasks, procedures, and stimuli. Thus, we could capture and better understand the effects of the evaluators and language of administration when assessing bilingual children.

Conclusions

This study offers a contribution to the fields of bilingual education, special education, and school psychology in the form of theoretical and practical implications of using qualitative assessments for accurately measuring bilingual children's cognitive and language development. Six patterns were supported by a critical review of the literature and qualitative analysis of data which are related to experts' suggestions for overcoming current methodological problems when assessing bilingual children. It is our argument that current assessment practices with bilingual children can be enhanced by attending to the theoretical and practical implications of the six proposed patterns which illustrate common dilemmas held by evaluators including: (1) misconceptions resulting from lack of awareness of the value of accurate qualitative assessments that measure the complex developmental process of becoming bicognitive, bicultural, and bilingual; (2) contradictory information obtained from qualitative and standardized assessments; and (3) the influence of evaluators' personalities on assessment.

In summary, we can conclude in light of findings from this study, that the particular cultural and linguistic content of qualitative assessments influences cognitive and linguistic performance in bilingual children, and that evaluators' personality, and not the instruments used, are the major assessment tools. Findings point to the need to conduct research for determining the influence of: (1) the relative proficiency in L1 and L2 on non-verbal conceptual development, (2) non-verbal abilities on L1 and L2 proficiency, and (3) individual differences on L1 and L2 proficiency. Finally, we can identify some methodological changes for improving future studies including the addition of standardized intelligence measures such as non-verbal tests and intelligence scales, as well as other qualitative measures such as teachers' ratings of first and second language.

Acknowledgments

The authors wish to thank Dr. Patricia Jones, Research Specialist at the Principal User Support Center for Computer and Information Technology at the University of Arizona for her time and valuable guidance in the middle stages of the non-parametric statistical analysis of data in this paper. The authors also want to express their appreciation to the fourteen graduate students majoring in educational psychology at the

University of Arizona who served as evaluators in partial fulfillment of the requirements for a course on testing of minorities that they took during Fall 1992.

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (1985). *Standards for educational and psychological testing*. Washington, DC: APA.
- Ballard, W. S., Tighe, P. L., & Dalton, E. F. (1979). *IDEA oral language proficiency test*. Whittier, CA: Ballard & Tighe Inc.
- Brown, L., Sherbenou, R. J., & Dollard, S. (1982). *Test of Nonverbal Intelligence (TONI)*. Austin, TX: Pro-Ed.
- Chamberlain, P., & Medinos-Landurand, P. (1991). Practical considerations for the assessment of LEP students with special needs. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 111-156). Austin, TX: Pro-Ed.
- Chapman, M. (1991). The epistemic triangle: Operative and communicative components of cognitive competence. In M. Chandler & M. Chapman (Eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities* (pp. 209-228). Hillsdale, NJ: Erlbaum.
- Commins, N. L. (1986). *A descriptive study of the linguistic abilities of four low achieving Hispanic bilingual students*. Unpublished Dissertation Abstract, Boulder, CO: University of Colorado.
- Cronbach, L. J. (1986). Signs of optimism for intelligence testing. *Educational Measurement Issues and Practices*, Fall, 23-24.
- Cummins, J. (1991). Inter-dependence of first-and- second-language proficiency in bilingual children. In Bialystock, E. (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge: Cambridge University Press.
- Frasier, M. M. (1991). Disadvantaged and culturally diverse gifted students. *Journal for the Education of the Gifted*, 14,(3), 234-245.
- Gonzalez, V. (1991). *A model of cognitive, cultural, and linguistic variables affecting bilingual Spanish/English children's development of concepts and language*. Doctoral Dissertation.

- Austin, Texas: The University of Texas at Austin. (ERIC Document Reproduction Service No. ED 345 562).
- Gonzalez, V. (1993a). Special voices. *NABE News*, 16(6), 23 & 26.
- Gonzalez, V. (1993b). Assessment of language-minority students: An awakening experience. *NABE News*, 17, (2), 9-10, & 26.
- Gonzalez, V. (1994). A model of cognitive, cultural, and linguistic variables affecting bilingual Spanish/English children's development of concepts and language. *Hispanic Journal of Behavioral Sciences*, 16 (4), 396-421.
- Gonzalez V. (1995). *Cognition, culture, and language in bilingual children: Conceptual and semantic development*. Bethesda, MD: Austin & Windfield.
- Gonzalez, V., Bauerle, P., & Felix-Holt, M. (1994a). A qualitative assessment method for accurately diagnosing bilingual gifted children. *NABE '92-'93 Annual Conference Journal*, 37-52. Washington, DC: NABE.
- Gonzalez, V., Bauerle, P., & Felix-Holt, M. (1994b). Deriving educational implications from using qualitative methods. *NABE News*, 17(7), 13,14, & 30.
- Gonzalez, V., & Yawkey, T. D. (1993). The assessment of culturally and linguistically diverse students: Celebrating change. *Educational Horizons*, 73(1), 41-49.
- Hakuta, K. (1988). Why bilinguals? In F. S. Kessel (Ed.). *The development of language and language researchers: Essays in honor of Roger Brown* (pp. 299-318). Hillsdale, NJ: Erlbaum.
- Hamayan, E. V., & Damico, J. S. (1991a). Developing and using a second language. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 39-75). Austin, TX: Pro-Ed.
- Hamayan, E. V., & Damico, J. S. (1991b). Implementing assessment in the real world. In E. V. Hamayan & J. S., Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 303-316). Austin, TX: Pro-Ed.
- Harley, B., Cummins, J., Swain, M., & Allen, P. (1990). The nature of language proficiency. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 7-25). Cambridge: Cambridge University Press.

- Kirsner, K., Smith, M., Lockhart, R., King, M., & Jam, M. (1984). The bilingual lexicon: Language specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, 23, 519-539.
- Lambert, W. E., Havelka, J., & Crosby, C. (1958). The influence of language acquisition context on bilingualism. *Journal of Abnormal Social Psychology*, 56, 239-244.
- Lewis, J. (1991). Innovative approaches in assessment. In R. J. Samuda, S. L. Kong, J. Cummins, J. Pascual-Leone, & J. Lewis (Eds.), *Assessment and placement of minority students* (pp. 123-142). Toronto: Hogrefe.
- Loyola, J. L., McBride, D. F., & Loyola, L. J. (1991). Cognitive and linguistic abilities of Puerto Rican bilingual children: Implications for assessment. *The Journal of Educational Issues of Language Minority Students*, 8, 31-50.
- Markman, E. M. (1984). The acquisition of hierarchical organization of categories by children. In C. Sophian (Ed.) *Origin in cognitive skills*. The 18th Annual Carnegie Symposium on Cognition (pp. 376-406). Hillsdale, N J: Erlbaum.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 2-11.
- Oller, J. W., Jr. (1991). *Language testing research: Lessons applied to LEP students and programs*. First Research Symposium on Limited English Proficient Students' Issues: Focus on Evaluation and Measurement. Washington DC: OBEMLA.
- Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of Verbal Learning and Verbal Behavior*, 34, 390-401.
- Piaget, J. (1967). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology, Vol. I* (pp. 703-732). New York: Wiley.
- Pollack, M. D. (1980). *The effects of testwiseness, language of test administration, and language competence on readiness test performance of low SES Spanish-speaking children*. University Microforms, International Microfiche N. 80-16031. Ann Arbor: University of Michigan.
- Potter, M., So, K., Von Eckardt, B., & Feldman, L. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning & Verbal Behavior*, 23, 23-38.

- Raven, J. C. (1976). *Coloured progressive matrices*. London: Lewis.
- Renzulli, J. S. (1991). The national research center on the gifted and talented: The dream, the design, and the destination. *Gifted Child Quarterly*, 35(2), 73-80.
- Reynolds, M. C., Wang, M. C., & Walberg, H. S. (1987). The necessary restructuring of special and regular education. *Exceptional Children*, 53, 391-398.
- Roth, J. (1988). Some assumptions about the evaluation of children. *Reading, Writing, and Learning Disabilities*, 4, 125-131.
- Samuda, R. J. (1991). Towards nondiscriminatory assessment: Principles and applications. In R. J. Samuda, S. L. Kong, J. Cummins, J. Pascual-Leone, & J. Lewis (Eds.), *Assessment and placement of minority students* (pp. 173-189). Toronto: Hogrefe.
- Sternberg, R. J. (1991). Death, taxes, and bad intelligence tests. *Intelligence*, 15, 257-269.
- United States General Accounting Office. (1987). *Case study evaluations*. Washington, DC: Program Evaluation and Methodology Division.
- Yin, R. K. (1994). *Case study research: Design and methods* (2nd Ed). Thousand Oaks, CA: Sage Publications.
- Waxman, S. R. (1990). Linking language and conceptual development: Linguistic cues and the construction of conceptual hierarchies. *The Genetic Epistemologist*, 17, 13-20.

Appendix

Table 1
*Frequencies of Conceptual Developmental Stages Attained by
 Bilingual Hispanic Kindergartners in Verbal Tasks of QUEST for
 General and Gender-Based Point Assignment Areas for Animal and
 Food Referents and for the Spanish and English Languages*

Verbal Tasks of Quest	Language of Administration			
	Spanish (n=11)		English (n=14)	
	Animals	Food	Animals	Food
Labeling General	8 high 8 med	9 high 2 med	9 high 5 med	4 high 9 med 1 low
Labeling Gender	11 low 13 low	11 low 13 low	1 high	1 med
Defining General	5 pre-op. 6 concrete	13 pre-op. 7 concr. 1 metalin.	5 pre-op. 7 concr. 2 metalin.	7 pre-op. 6 concr. 1 metalin.
Defining Gender	11 pre-op	11 pre-op.	12 pre-op. 2 concr.	13 pre-op 1 concr.
Verbal Just. of Sorting General	2 pre-op. 18 concr. 1 metalin.	1 pre-op. 4 concr. 6 metalin.	2 pre-op. 11 concr. 1 metalin.	8 pre-op. 14 concr.
Verbal Just. of Sorting Gender	2 pre-op. 9 concr.	5 pre-op. 1 concr. 5 metalin.	6 pre-op. 8 concr.	12 pre-op. 2 concr.
Category Clue Verbal Just. General	2 pre-op. 9 concr.	5 pre-op. 4 concr. 2 metalin.	6 pre-op. 8 concr.	9 pre-op. 5 concr.
Category Clue Verbal Just. Gender	3 pre-op. 8 concr.	7 pre-op. 1 concr. 3 metalin.	9 pre-op. 5 concr.	13 pre-op. 1 concr.

Table 2
Frequencies of Conceptual Developmental Stages Attained by Bilingual Hispanic Kindergartners in Non-Verbal Tasks of QUEST for General and Gender-Based Point Assignment Areas for Animal and Food Referents and for the Spanish and English Languages

Non-verbal Tasks of Quest	Language of Administration			
	Spanish (n=11)		English (n=14)	
	Animals	Food	Animals	Food
Sorting General	1 pre-op. 10 concr.	1 pre-op. 7 concr. 3 metalin.	1 pre-op. 13 concr.	5 pre-op. 9 concr.
Sorting Gender	3 pre-op. 3 concr.	7 pre-op. 1 concr. 3 metalin.	6 pre-op. 8 concr.	12 pre-op. 2 concr.
Category Clue Sorting General	1 pre-op. 10 concr.	2 pre-op. 5 concr. 4 metalin.	3 pre-op. 11 concr	6 pre-op. 7 concr. 1 metalin.
Category Clue Sorting Gender	2 pre-op. 9 concr.	6 pre-op. 1 concr. 4 metalin.	7 pre-op. 7 concr.	13 pre-op. 1 metalin.

Table 3
Summary of X2 Tests of Association, Degrees of Freedom, Cramer Vc Statistic, and Level of Significance for the Two-Way Tables for Animal and Food Referents, and for the Spanish and English Languages

Language	Tasks	Referents	Chi Square	Cramer Vc Statistic	d	p
----------	-------	-----------	------------	---------------------	---	---

First set: Conceptual development achieved for general and gender-based areas of point assignments versus conceptual development achieved for S, VJS, CCS, and CCVJS tasks.

Spanish	CC. vs. S	Animal	4.950	.671	1	.05
Spanish	CC. vs. VJS	Animal	6.519	.770	1	.05
	CC. vs. VJS	Animal	5.833	.645	1	.05
English	CC. vs. S	Food	14.00	1.000	2	.01

Second set: Conceptual development achieved for non-verbal tasks (S and CCS) versus conceptual development achieved for verbal tasks (VJS and CCVJS).

General areas of point assignment

Spanish	CCS vs. CCVJS	Food	10.340	.686	4	.05
English	S vs. VJS	Animal	6.462	.679	2	.05
	CCS vs. CCVJS		5.091	.603	1	.05
English	S vs. VJS	Food	5.833	.645	1	.05
	CCS vs. CCVJS		6.533	.683	1	.05

Gender-based areas of point assignment

Spanish	S vs. VJS	Animals	6.519	.770	1	.05
Spanish	S vs. VJS	Food	15.714	.845	4	.01
	CCS vs. CCVJS		18.071	.906	4	.01
English	S vs. VJS	Animals	7.024	.708	1	.01
	CCS vs. CCVJS		7.778	.745	1	.01
English	S vs. VJS	Food	14.000	1.000	1	.001
	CCS vs. CCVJS		14.000	1.000	1	.001

Third set: Conceptual development achieved for animal and food referents versus conceptual development achieved for L, D, S, VJS, CCS, and CCVJS.

Gender-based areas of point assignment

English	Labeling	Animals vs. Food	14.000	1.000	1	.001
English	Defining	Animals vs. Food	6.462	.679	1	.5

Note: L= Labeling, D= Defining, S= Sorting, VJS= Verbal Justification of Sorting, CCS= Category Clue Sorting, and CCVJS= Category Clue Verbal Justification of Sorting.