Potthoff Bias Analyses of K-ABC MPC and Nonverbal Scale IQs Among Anglo, Black, and Puerto Rican Kindergarten Children

Joseph J. Glutting
University of Delaware

Bias analyses, via Potthoff's (1966) regression technique, were conducted to investigate the predictive utility of Kaufman Assessment Battery for Children (K-ABC) Mental Processing Composite (MPC) and Nonverbal Scale IQs with a random sample of 146 Anglo, Black, English-dominant Puerto Rican, and Spanish-dominant Puerto Rican kindergartners children on four achievement criteria (K-ABC Achievement Scale, Needs Assessment Survey Reading and Mathematics subscales, and Teacher-Assigned Grades). Bivariate correlations showed that Nonverbal Scale IQs were poor predictors for English-dominant Puerto Rican children. However, only one of the 15 Potthoff comparisons showed a group difference toward biasing the predictors against the minority children: MPC slope and intercept differences underestimated low-scoring Blacks on the Grade criterion. When significant differences were found during other comparisons, results consistently showed that K-ABC IQs overestimated achievement for the minority students. Results are discussed in terms of the implications predictive bias and predictive validity issues have for school psychologists.

Legislative mandates and litigious child advocacy have thrust test fairness and test bias into the lexicon of school psychologists. The issue of test fairness pertains to the decision model used to determine program admission or service eligibility. Currently, society leans toward using the "representative equality" model of selection (see Nichols, 1978, for a discussion of other selection models). Under this model, program participation is established in proportion to a group's numerical representation in the population. Test bias, on the other hand, is a quantitative, statistical term denoting the presence of constant or systematic error between majority and minority-group performances on an instrument.

The administration of an unbiased test does not ensure fairness under the representative equality model because an unbiased measure may continue to over- or under-select groups for program admission. Although a matter of consequence, the numerical representation of subgroups in educational programs is a social issue. Reduced bias, on the other hand, is an empirical matter that involves the examination of a test's validity. Reduced bias also is a necessary requisite to test fairness because the relationship between bias and fairness is similar to the linkage between test reliability and validity (Reynolds &
Brown, 1984). That is, as reliability is a requisite for test validity, so too must a test be unbiased in order to be capable of being fair. Like validity paradigms, tests can be studied from the perspectives of content, construct, and predictive bias. School psychologists must consider all three sources of bias when choosing instruments for use with children. However, "from the standpoint of traditional applications of aptitude and intelligence tests in forecasting probabilities of future performance levels, predictive validity is the most crucial form of validity in relation to bias" (Reynolds, 1984, p. 28).

Predictive bias with individually administered intelligence tests rarely has been found, and when significant findings have been obtained, results have tended to show higher validity for minority than majority-group children (see Jensen, 1980; Reynolds, 1982a, 1982b, for reviews). Even so, research generally has been limited to variations between Anglo-Black, or Anglo-Mexican-American groups. Far fewer studies have been conducted with other minority cohorts, and there has been even less effort at the preschool level where diagnostic decisions have a longer-lasting impact on children (Mitchell, 1967; Oakland, 1978; Reynolds, 1980, 1983; Reynolds & Pielser, 1983; Reynolds, Williams, & Chatman, 1985).

Another concern is the extent to which test scores predict classroom performance. Predictive bias studies have been criticized for employing achievement tests as the sole criterion of academic success (Hilliard, 1984; Kamin, 1974; Mercer, 1979, 1984). Teacher ratings and classroom grades have been eschewed by researchers because these measures have insufficient scope for a range of the increasing homogeneity of ability levels found at each ensuing grade. Nonetheless, tests must do more than correlate with one another in grades where student heterogeneity is sufficient to provide adequate score ranges, with valid tests being measures that predict scholastic performance in actual classroom situations. Much of the speculation surrounding the recent publication of the Kaufman Assessment Battery for Children (K-ABC) has centered on the instrument's claim of reduced Anglo-Black score differences. Although well intended, discussions of majority-minority mean discrepancies can be misleading because mean score differences have rarely been used as a standard for determining bias (Anastasi, 1979; Bracken, 1985; Clarizio, 1982; Sattler, 1982). The more common benchmark for estimating utility is to examine an instrument's proficiency in making predictions with the least feasible random error, and to evaluate whether there is constant error, among majority and minority groups (Cleary, Hamphreys, Kendrick, & Wexman, 1975).

More simply stated, "a test is not biased if individuals from different groups have the same test scores also have the same expected criterion score" (Pothoff, 1966, p. 7). Pothoff's definition not only is straightforward and in accord with other widely accepted interpretations of predictive bias (Cardall & Coffman, 1964; Cleary, 1968; Cleary et al., 1975; Reynolds, 1982b), the definition permits an empirical test of the issue through regression analyses. According to Pothoff (1966) then, bias in prediction is a function of slope and/or intercept differences among (a) the groups studied, (b) the specific test scores employed, and (c) the criteria examined.

The purpose of the study is to investigate the predictive utility of K-ABC Mental Processing Composite, and K-ABC Nonverbal Scale IQs at the kindergarten level. Specifically, the study examines are IQ slope and intercept differences among four ethnic groups (Anglo, Black, English-dominant Puerto Rican, and Spanish-dominant Puerto Rican) on one test (K-ABC Achievement Scale) and three teacher-rated criteria of achievement obtained at different intervals during the same school year (Needs Assessment Survey Reading and Mathematics Scales, and Teacher-Assigned Grades).

**Subjects**

A random sample of 146 children, stratified by ethnicity, was obtained from a public school population of 550 students attending kindergarten in a combination urban-rural school system located on the east coast of the United States. All of the children were attending kindergarten for the first time. The students ranged in age from 5 years, 0 months through 6 years, 11 months. Retained children from the previous school year and students below the age of 5 were excluded from the study in an effort to assure age-range homogeneity. The mean age for the sample was 65.5 months with a standard deviation of 3.57 months. There were 79 boys (54%) and 67 girls (46%).

The system where the study was conducted is ethnically diverse: 65 (45%) of the children were Anglo, 24 (16%) were Black, and 57 (39%) were Puerto Rican. A three-step procedure was used to determine the language preference of Puerto Rican students. Subdividing the Puerto Rican group, the proficiency was assessed by research which has shown that assessment errors occur when the language background of Puerto Rican preschoolers is ignored (Glutting, 1984, 1985; Glutting, Baker, & Gehard, 1983). First, all parents completed a home language survey when they registered their children for school. Children whose parents indicated that they speak Spanish at home were administered the Language Assessment Battery Level K-2 (LAB; Board of Education of the City of New York, 1976) upon school entrance. Children who met 17 of 29 LAB criteria for Spanish dominance were then placed in self-contained bilingual classrooms for a trial period. Within 2 weeks, the children were evaluated on the Language Assessment Scale Level K-5 (LAS; DeAville & Duncan, 1977). A child was placed in the Spanish-dominant group, and remained in a self-contained bilingual classroom, when 47 out of a possible 50 LAB criteria showed Spanish dominance, 27 (18% of the total sample) of the Puerto Rican children were placed in the English-dominant Puerto Rican group, and 30 (21%) were placed in the Spanish-dominant Puerto Rican group.

**Instruments**

**Predictors**

Kaufman Assessment Battery for Children: Mental Processing Composite (Kaufman & Kaufman, 1983a, 1983b). The K-ABC is composed of separate, but related, tests of intelligence and achievement designed for the evaluation of children aged 2½ to 12½. The battery yields four global scores. Two of the scales, Sequential Processing Scale, and Simultaneous Processing Scale, have separate IQs from select subtests within the Successive and Simultaneous Processing Scales. Although all of the mental-processing subtests were designed to minimize the effect of language on performance, the Nonverbal Scale was developed for the assessment of hearing-impaired, language-disordered, and non-English-speaking children ages 4 through 12½ (Kaufman & Kaufman, 1983a).
Nonverbal IQ is an appropriate predictor for all of the demographic groups in the study, including the Spanish-dominant Puerto Rican children.

Criteria

Kaufman Assessment Battery for Children: Achievement Scale (Kaufman & Kaufman, 1983a, 1983b). The Achievement Scale of the K-ABC is divided into six subtests, four of which are appropriate for use at the kindergarten level (Faces and Places, Arithmetic, Riddles, and Reading/Decoding). A global achievement score is obtained by adding subtest standard scores by using a conversion table in the K-ABC manual.

Needs Assessment Survey (NAS). The NAS is an academic rating scale completed by a child's classroom teacher and is appropriate for use in Grades 1-4. The survey form, developed by the district where the study took place, is divided into twelve scales with fifteen items per area. The Reading Scale (NASR) includes items that evaluate listening, auditory comprehension, visual discrimination, and basic sight vocabulary, and so on, whereas the Mathematics Scale (NASM) contains items for the memorization of addition facts, process understanding for addition, and such. Teachers rate students on each item using a 4-point index ranging from extremely great need for help (1) to no need for help (4) with raw scores ranging from 0 to 60 on each of the two measures. Previous research with the instrument has shown significant concurrent first-grade correlations with the Metropolitan Achievement Test Reading and Mathematics subtests for both the NASR (r = .72 and .56, respectively) and NASM (r = .31 and .51, respectively) (Glutting, 1984).

Teacher-Assigned Grades. Report-card marks are the cornerstone by which parents judge academic progress. Consequently, end-of-the-year report-card grades were included as the third criterion. Nine academic areas were rated on the report cards (e.g., recognizes and names colors, and recognizes and numbers numbers) with raw scores ranging from 0 through 27.

Procedure

The investigation was delayed until mid-October to allow the children time to acclimate to school. The K-ABC battery was then administered individually until mid-February of the children's kindergarten year. Reactivity to the assessment process was reduced by including the tests as part of a regularly occurring screening program. All demographic groups were assessed on the K-ABC across each 1-week period during the study.

Three Anglo doctoral candidates in school psychology administered the K-ABC battery in English to the Anglo, Black, and English-dominant Puerto Rican groups. Following Kaufman's (personal communication, March 14, 1983) recommendations, all assessments with the Spanish-dominant Puerto Rican children were limited to standardized pan-dominated administrations of the Nonverbal Scale subtests. As a result, MPC and K-ABC Achievement Scale scores were not computed for the Spanish-dominant Puerto Rican group.

During the last week in May, teachers completed the Reading and Mathematics sections of the Needs Assessment Survey for the children's dominant language. Teacher-assigned grades also were obtained for the dominant language of the students. All grades came from the June marking period, which was over 4 months after the last administration of the K-ABC.

Data Analysis

Pearson correlations were calculated using program Multiple Correlation and Regression (Cron, 1985) and Potthoff analyses were conducted using program Micro Congru (Watkins, 1980). An advantage of Potthoff's (1966) technique over other bias procedures is that it reduces the Type I error rate by providing a "Simultaneous Test" for the equivalence of both regression coefficients and intercepts between the groups under study (Reynolds, 1982a). A significant F shows that bias exists and signifies that separate slope and intercept analyses should be conducted to determine the type(s) of bias present. When either a significant slope or intercept difference is found between two groups, the cohort with the higher mean criterion score is underscored. The situation is compounded though when both slope and intercept differences are found. Here, group members with the lower mean criterion score typically are underpredicted when they obtain low scores on the test under study, but their criterion scores are overestimated when they obtain high scores on the test.

Results

Internal consistency was estimated for the NASR and NASM to evaluate their reliability as criteria. Alpha coefficients (Cronbach, 1951) indicated that the instruments are reliable measures of reading (r = .91) and mathematics achievement (r = .93). No internal consistency coefficients were calculated for the K-ABC Achievement Scale criterion because of the heterogeneity of its item content, or for the Grade criterion because of its uniform acceptance as a measure of academic performance.

Standard scores from the two K-ABC IQs (MPC and Nonverbal IQ) and the K-ABC Achievement Scale were used during the regression analyses along with raw scores from the three other measures (NASR, NASM, and Grade criteria). Table 1 presents the means for the four ethnic groups on the predictors and criteria.

As previously indicated, the MPC and K-ABC Achievement Scale were not administered to the Spanish-dominant Puerto Rican group because of their limited English proficiency. Predictive coefficients between the criteria and the MPC are presented at the top of Table 2 for the Anglo, Black, and English-dominant Puerto Rican groups. An overall coefficient also is presented for the combined MPC scores from the three

<table>
<thead>
<tr>
<th>Predictr</th>
<th>Ethic Group</th>
<th>English-Dominant Puerto Rican</th>
<th>Spanish-Dominant Puerto Rican</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-ABC MPC</td>
<td>102.91</td>
<td>89.96</td>
<td>93.22</td>
</tr>
<tr>
<td>K-ABC Nonverbal IQ</td>
<td>104.00</td>
<td>89.83</td>
<td>98.00</td>
</tr>
<tr>
<td>Criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASR</td>
<td>52.29</td>
<td>44.46</td>
<td>46.78</td>
</tr>
<tr>
<td>NASM</td>
<td>53.32</td>
<td>48.23</td>
<td>50.44</td>
</tr>
<tr>
<td>GRADE</td>
<td>21.75</td>
<td>21.50</td>
<td>23.82</td>
</tr>
<tr>
<td>K-ABC Achievemnt Scale</td>
<td>96.71</td>
<td>84.88</td>
<td>84.07</td>
</tr>
</tbody>
</table>

Note. The K-ABC MPC and Achievement Scale were administered in English and were not given to the Spanish-dominant Puerto Rican group.
The results show that Black kindergartners with low MPC scores were underpredicted on the Grade criterion, whereas high-scoring Blacks were overpredicted.

Potthoff (1966) analyses with the K-ABC Nonverbal IQ were conducted using all four ethnic groups on the NASM (see the bottom of Table 3). The English-dominant Puerto Rican group was excluded, however, from the NASR and Grade analyses because of the nonsignificant bivariate correlations found between their Nonverbal IQs and the two criteria. The analyses on the NASR and NASM therefore were limited to comparisons among the Anglo, Black, and Spanish-dominant Puerto Rican groups. Significant simultaneous test results were found on all three criteria after the application of Bonferroni adjustments. Nonetheless, separate slope and intercept analyses among the groups showed only slope differences on the NASM and Grade criteria and only intercept differences on the NASR, with Anglos obtaining the highest mean on each criterion (see Table 1). The results therefore show that the K-ABC Nonverbal IQ is either unbiased, or overpredicts, achievement on the NASR, NASM, and Grade criteria for Black and Spanish-dominant Puerto Rican children and that the Nonverbal IQ is an unbiased predictor of NASM ratings for the English-dominant Puerto Rican group.

### Tables

#### Table 2

<table>
<thead>
<tr>
<th>Group</th>
<th>NASR</th>
<th>NASM</th>
<th>Grade</th>
<th>K-ABC Achievement Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anglos</td>
<td>.66</td>
<td>.63</td>
<td>.55</td>
<td>.71</td>
</tr>
<tr>
<td>Blacks</td>
<td>.65</td>
<td>.75</td>
<td>.76</td>
<td>.68</td>
</tr>
<tr>
<td>English Dominant Puerto Ricans</td>
<td>.64</td>
<td>.60</td>
<td>.39</td>
<td>.73</td>
</tr>
<tr>
<td>All Three Ethnic Groups</td>
<td>.65</td>
<td>.67</td>
<td>.61</td>
<td>.73</td>
</tr>
</tbody>
</table>

Note. MPC = Criteria coefficients were not calculated for the Spanish-dominant Puerto Rican group.

#### Table 3

<table>
<thead>
<tr>
<th>MPC—Criteria Comparisons</th>
<th>Overall Analysis</th>
<th>Simultaneous Test</th>
<th>Slope Test</th>
<th>Intercept Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NASR</td>
<td>NS</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>NASM</td>
<td>NS</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>Grade</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonverbal IQ—Criteria Comparison</th>
<th>Simultaneous Test</th>
<th>Slope Test</th>
<th>Intercept Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p &lt; .01</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

*Analyses were conducted among the Anglo, Black, and English-dominant Puerto Rican groups.

### Discussion

School psychologists need to be aware that the evaluative process becomes valid only when classifications prescribe treatment or predict what will happen when effective treatments are unavailable (Gough, 1971). Implicit here is that test scores become meaningful when they predict future functioning or prescribe remedial procedures for children with exceptional characteristics. Traditional IQ tests have been shown to offer little in the way of prescriptively relevant information (for a review, see Ysseldyke & Milch, 1982). Therefore, unless prescriptive relevance is satisfactorily demonstrated for the processing scores in the K-ABC, the practical utility of its IQs will rest on their ability to estimate criterion performance with the least feasible random error among majority and minority groups.

### Table 2

Bivariate Correlations Between the MPC and Nonverbal Scale IQ of the K-ABC and the Four Achievement Criteria

<table>
<thead>
<tr>
<th>Group</th>
<th>MPC—Criteria Coefficients</th>
<th>K-ABC Achievement Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anglos</td>
<td>.66</td>
<td>.63</td>
</tr>
<tr>
<td>Blacks</td>
<td>.65</td>
<td>.75</td>
</tr>
<tr>
<td>English Dominant Puerto Ricans</td>
<td>.64</td>
<td>.60</td>
</tr>
<tr>
<td>All Three Ethnic Groups</td>
<td>.65</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note. MPC = Criteria coefficients were not calculated for the Spanish-dominant Puerto Rican group.

#### Table 3

Potthoff Analyses Between the MPC and Nonverbal Scale IQ of the K-ABC and the Four Achievement Criteria

<table>
<thead>
<tr>
<th>MPC—Criteria Comparisons</th>
<th>Overall Analysis</th>
<th>Simultaneous Test</th>
<th>Slope Test</th>
<th>Intercept Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NASR</td>
<td>NS</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>NASM</td>
<td>NS</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>Grade</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonverbal IQ—Criteria Comparison</th>
<th>Simultaneous Test</th>
<th>Slope Test</th>
<th>Intercept Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p &lt; .01</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

*Analyses were conducted among the Anglo, Black, and English-dominant Puerto Rican groups.

*Analyses were conducted among the Anglo, Black, English-dominant Puerto Rican, and Spanish-dominant Puerto Rican groups.
Only one of the 15 Potthoff comparisons in the current study showed a significant difference in the direction of biasing predictions against the minority children; MPC slope and intercept differences were biased against low-scoring Blacks on the Grade criterion. When significant differences were found during the other comparisons, findings were uniform in showing that K-ABC IQs overestimated achievement for the minority groups. Consequently, the Potthoff results are in accord with previous preschool research (Reynolds, 1980, 1983; Reynolds et al., 1985) because the relationship between the MPC and achievement, and the Nonverbal Scale IQ and achievement, generally was independent of ethnic-group membership for the Anglo, Black, English-dominant Puerto Rican, and Spanish-dominant Puerto Rican students.

The Potthoff findings also reveal that K-ABC IQs were valid predictors of classroom performance. Results from the slope and intercept analyses showed little evidence of bias on the NASM, NASR, or Grade criteria. Thus, the K-ABC IQs obtained at the kindergarten level were found to have pragmatic validity for the real-world criteria of classroom performance. The findings indicate, then, that school psychologists can use K-ABC IQs with kindergarten children to predict both achievement-test scores and the more relevant criterion of classroom success.

Clinical tradition maintains that validity is enhanced when nonverbal tests are used with Spanish-speaking children. Bivariate correlations showed that MPC and Nonverbal Scale IQs from the K-ABC provide meaningful estimates of achievement for the ethnic groups studied. The one exception occurred with the Nonverbal IQs for the English-dominant Puerto Rican group; only one of the three Nonverbal Scale validity coefficients was significant (NASM), while all four MPC-achievement correlations (NASM, NASR, Grade, and K-ABC Achievement Scale) exceeded chance. Contrary to popular belief, the results show that school psychologists actually maximize the accuracy of kindergarten-level predictions when they use the more language-loaded MPC scores from the K-ABC, and ignore Nonverbal Scale results, when estimating achievement for English-dominant Puerto Rican students.

Preschool assessment has far-reaching implications for the lives of children and their families. As a result, preschool instruments must be held to the same rigid standards imposed on the development and use of other tests, with validity becoming the most important consideration in test evaluation (cf. Standards for Educational and Psychological Testing, American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985). The current findings indicate that K-ABC Nonverbal Scale IQs meet concurrent and predictive validity requirements for Spanish-dominant Puerto Rican children at the kindergarten level. Nonverbal Scale scores for this group were found to be unbiased predictors and showed significant and substantial correlations with achievement across a 4-month period. The results therefore reduce some of the uncertainty school psychologists face when evaluating Spanish-dominant Puerto Rican kindergarten by showing that they can place confidence in the short-term predictions provided by the Nonverbal Scale IQ in the K-ABC.

ACKNOWLEDGMENTS
I wish to express appreciation to Carl Simmons, Bernice Steibert, and Margaret Bitters of the Vineland Public Schools for their assistance during the investigation.

REFERENCES
Methods of Skill Selection in Social Skills Training: A Review

Jan N. Hughes
Texas A&M University

Thirty-two social skills training studies, utilizing a social learning or social-cognitive approach and published between 1980 and 1994, were reviewed with respect to methods used to select skills targeted for training. Operational definitions of social skills selection methods were used to classify studies along two dimensions: type of social validity and degree of individualization. Social validity refers to selecting skills based on previous research that has established a relation between the targeted skill and socially important outcomes. Individualization refers to selecting skills based on an evaluation of trainers’ preservice performance. Judges obtained 94% agreement in applying these operational definitions to classify studies. A gap between recommended methods for selecting skills for inclusion in social skills training programs and actual research practices was found. Of the 32 studies reviewed, 18 (56%) reported no prior empirical evidence that the targeted skills were socially valid, and 20 (63%) did not report an attempt to verify that trainers were deficient in the targeted skills prior to treatment. Implications of these findings for researchers and clinicians in school psychology are discussed.

The past 10 years have seen an increase in journal articles and books devoted to the topic of children’s social skills. Earlier efforts to enhance children’s social skills employed either contingent reinforcement and punishment strategies (e.g., Becker, Madsen, Arnold, & Thomas, 1967; Walters, Parke, & Case, 1965; Ward & Baker, 1960) or modeling (O’Connor, 1969, 1972). More recent efforts (e.g., Gresham & Nagle, 1980; Ladd, 1981) represent diverse strategies, including instruction, modeling, rehearsal of specific behaviors, performance feedback, and social-cognitive problem-solving training. These recent efforts have broader, more global goals and attempt to effect positive change in a child’s social competence rather than only a change in specific behaviors. For example, in Ladd’s (1981) study, children were instructed in, among other skills, offering supportive statements to peers, and dependent measures included peer sociometrics. Although the terms social competence and social skills are often used interchangeably, social competence is a more general term, reflecting overall adequacy of social skills, and social skills refer to the component behavioral or cognitive skills leading to social competence (Kraitochwill & French, 1984). In this regard, peer acceptance and friendships are the most frequently recommended measures of social competence (Asher & Hymel, 1981; Ladd, 1981; Pataalla & Gottman, 1983), and behavioral observations or checklists are used for the assessment of such particular social behaviors such as expressing a compliment or offering assistance to a peer (Asher & Hymel, 1981; Michelson, Suga, Wood, & Kazdin, 1983). Adding to the complexity is the fact that some definitions of social skills