Abstract: In this study, predictive classification accuracy was used to select those tasks from a kindergarten screening battery that best identified children who, three years later, were classified as educable mentally handicapped or as having a specific learning disability. A subset of measures enabled correct classification of 91% of the children in those special education categories and 85% of the children in regular education. Although there was a high percentage of false positives, 57% of these children were actually poor achievers in third grade. Lower screening scores were associated with being tested in Spanish or both English and Spanish, minority status and low SES.

More than a decade ago, Public Law 99-457 (1986) mandated the provision of public school education for preschool children with handicaps. This legislative act necessitated identification of all preschool children eligible for publicly funded early specialized educational programs. Of all those who are eligible, the group most difficult to identify at an early age are those with mild learning problems (Jacob, Snider, & Wilson, 1988; Meisels, 1989; Mercer, Algozzine, & Trifiletti, 1988; van Kraayenoord, 1983). They are identified later in grade school, even when an extensive screening program is in place (Scott, Urbano, & Boussey, 1991). Children classified as educable mentally handicapped (EMH) or as having a learning disability (LD) are examples of two categories of special needs children typically described as having mild learning problems (Gresham, MacMillan, & Bocian, 1996; Holtzman & Wilkinson, 1991).

Clearly, full implementation of PL 99-457, now called the Individuals With Disabilities Education Act, will not be possible until screening tests are developed that more effectively identify young children with mild learning problems. Use of such screening measures presupposes that the “… behavior and cognition of young children is malleable and that early intervention will facilitate appropriate development” (Salvia & Ysseldyke, 1991, p. 467). Consistent with this belief, early intervention has proven to be a) more effective for those who participated from an early age (Shonkoff & Hauser-Cram, 1987), b) associated with a reduced need for subsequent special education (Lazar & Darlington, 1982), and c) more effective for children with mild delays (Dunst, Trivette, & Cross, 1986).

As one potential remedy for the difficulties encountered in the early detection of children with mild learning problems, Scott and her colleagues (Scott, Greenfield, & Sterental, 1986) have been designing and evaluating new cognitive screening tests that are composed of simple tasks, but tasks that require the active engagement of children’s cognitive processing skills (Greenfield & Scott, 1985).
For example, children might be shown a picture consisting of many cats and be asked to tell what they look like and what they do (Scott, Fletcher, Jean-Francois, Urbano, & Sanchez, 1998) or they might be shown several arrays of three, nine-square matrices, two identical and one different pattern of colored dots displayed on each three-matrix array, and be required to point to the matrix with the different design (Scott, Deuel, Urbano, Fletcher, & Torres, 1998). To respond correctly, children must activate their cognitive processing capabilities and analyze the stimulus problem. A simple automatic associative response will not do. The use of their cognitive processing capabilities was expected to make overt, their not-yet-identified mild cognitive deficiencies.

Use of cognitive tasks is consistent with the finding that cognition is one of the developmental areas related to subsequent educational status (Lichtenstein & Ireton, 1991) and later school achievement (Funk, Sturner, & Green, 1986). Additionally, cognitive capabilities assessed prior to, or at the beginning of kindergarten have been shown to relate to both reading and math performance (e.g., Mantzicopoulos & Morrison, 1994; Swanson, 1994; Tramontana, Hooper, & Selzer, 1988). These are major areas of learning difficulty for children classified as EMH or LD, the latter condition now called specific learning disability (SLD), which more clearly reflects the limited extent of the problems typically observed with these children.

Following evaluations of individual tasks (e.g., Scott, Perou, Greenfield, & Swanson, 1993) and then batteries of multiple tasks (e.g., Scott, Deuel, Claussen, & Sanchez, 1993), a series of three concurrent validity studies was conducted with preschool children in which several versions of a screening battery were evaluated (e.g., Scott, Deuel, et al., 1998). For each version, the accuracy with which preschool children already classified as EMH or LD, the latter condition now called specific learning disability (SLD), which more clearly reflects the limited extent of the problems typically observed with these children.

We first examined the predictive capabilities of the screening performance of the preschool sample. The results were encouraging. We found that four years later, a set of just three tasks from the battery accurately predicted the third grade educational placement of 79% of the children with an EMH or SLD label and 70% of the children who were in regular education (Scott & Delgado, 2003). In this study, predictive accuracy of the screening performance of the kindergarten sample was evaluated. Once again, their educational status in third grade was the primary outcome measure.

Given previous results in both concurrent (e.g., Scott, Deuel, et al., 1998; Scott, Fletcher,
Jean-Francois, et al., 1998) and predictive (Scott & Delgado, 2003) studies, as well as data from other researchers (e.g., McIntosh, 1999; Stebbins & McIntosh, 1996), it was expected that the most accurate predictions would be associated with a subset of tasks from the screening battery, not with the total screening score.

Minority status (Scott et al., 2003; Scott, Fletcher, Jean-Francois, et al., 1998) and testing in Spanish or both Spanish and English (Scott et al., 2003) were associated with lower screening scores. In the preschool predictive study (Scott & Delgado, 2003), testing in Spanish or both languages was shown to be confounded with low SES, which in turn was related to poor screening performance and risk for poor long-term academic outcome. The confounding of minority status with low SES is consistent with other findings (e.g., Halle, Kurtz-Costes, & Mahoney, 1997) as is the relation between poor academic performance and low SES (Hill, 2001). The relation of testing language, minority status and SES to screening performance and academic achievement were further explored in the present study.

Method

Participants

Kindergarten sample. Four hundred and fifty-nine kindergarten children were administered a screening battery consisting of nine cognitive tasks in the 1993-94 school year. This sample was composed of every child from the targeted kindergarten classes in the 35 participating public schools for whom an informed consent was obtained. Informed consent included permission to access the child’s educational status and SAT scores through third grade. All six regions of the Miami/Dade County Public School (M/DCPS) system were represented. According to parental designations, 39% of the sample were classified as Black, 34.5% as Hispanic, and 26.5% as White. These three racial/ethnic classifications were the only ones used by the school system at that time. Clearly minority children make up most (73.5%) of the kindergarten sample. Fifty-nine percent of the sample was male and the

mean chronological age was 72.0 months ($SD = 4.2$).

Third grade predictive sample. Of the 459 kindergarten children with screening scores, 356 were successfully tracked into third grade. Of these, 345 were in regular education and 11 were in special education with either an EMH (1) or SLD (10) classification.

According to statewide criteria, for a child to be classified as EMH, he or she must achieve a full scale IQ score that falls between 2 and 3 $SD$s below the mean on two measures of intelligence and achieve a score on a test of adaptive behavior such as the Scales of Independent Behavior-Revised (Bruininks, Woodcock, Weatherman, & Hill, 1996) or the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) that falls at least 1.5 $SD$s below the mean of a child of the same age and socio-cultural group.

According to the same state wide criteria, for a child to be classified as LD, there must be a) a significant difference between language performance and other developmental levels, or b) a significant difference between receptive and expressive language abilities, or c) a significant language delay based on criteria presented in the test or evaluation manual.

All six regions in the M/DCPS system were still represented in the third grade predictive sample. The racial/ethnic distribution of this smaller sample was 41% Black, 35% Hispanic and 24% White. Fifty-six percent were male. Mean chronological age of this sample when the children were in kindergarten was 72.0 ($SD = 4.3$). This primarily minority sample (76%) is nearly identical to the kindergarten sample in terms of race/ethnicity and also mimicked the total school population which was approximately 83% minority in the 93/94 school year, the year the kindergarten children were administered the screening battery.

Kindergarten children excluded from the predictive sample. There were 101 children not included in the predictive sample. Seven were excluded because they were in a grade other than third, the grade they should have been in during the 1996-97 school year, and the reasons for this were not known. Twenty-seven children were never found in the school records after kindergarten and 57 children were not present in the third grade records although they had been found in earlier
grades. An additional 12 children were excluded because they were in special education classifications other than EMH or SLD.

There were more White children (35%), fewer Black children (33%), and more males (67%) in the group of children not included in the predictive sample. However, the mean chronological age of this excluded group was 72.0 (SD = 3.9) which was the same as mean chronological age of the 356 children in the predictive sample.

Testing Language

If a child lived in a home in which a language other than English was spoken, that child was tested in the language in which the child was most proficient. This was determined through a consideration of teacher nomination, the child’s choice and the child’s conversational ability. Spanish was the only other language used for testing. If the examiner was not certain that the child understood the instructions in the language employed, the examiner administered instructions in both languages. Of the 125 children in the third grade predictive sample designated Hispanic, 81% chose to be tested in English. Fifteen percent chose to be tested in Spanish and the screening battery was administered in both languages to only 4%. Examiners fluent in Spanish translated instructions into that language for those children who needed to be tested in Spanish or both languages. Responses in Spanish were accepted from any child regardless of testing language.

Screening Battery Contents and Administration

The screening battery consisted of nine tasks. Test items included colored photographs of meaningful pictures, or in one case colored dots. These stimuli were placed on 35.6 cm wide by 21.6 cm high white paper pages, which were placed in a four-hole legal size black binder, 36.8 cm wide by 25.4 cm high. The tasks are briefly described below in the same order as they appeared in the screening battery. For more details about the tasks and scoring see Scott, Deuel, et al. (1998).

**Picture pointing.** Six pictures were displayed on each of four pages. Children were asked to point to each picture just one time. The score for each array ranged from 0 (omission or repetition present) to 3 (pointing in a reading order sequence). Two dependent measures were evaluated: a quality score, which was the sum of the scores over the four pages, and a penalty score which was the quality score minus the total number of omissions and repetitions.

**Picture recognition.** After presenting a training memory page and test page, children were shown two sets. Each set consisted of one memory page and two recognition test pages. The examiner named each of the eight pictures on the memory page twice. On each of the two recognition test pages there were four “seen before” pictures and four new pictures. As test pages were presented, children were asked to point to the pictures that were the same as those from the memory page. The dependent measure was number of correct selections minus number of incorrect selections summed over the two sets.

**Word meaning.** Children were asked, “What is an airplane?” and then “What is a banana?” Scores for each part of the definition varied from 0 (statement not true) to 3 (abstract categorical).

**Standard oddity.** Children were asked to point to the different picture in each of nine unique arrays consisting of two identical and one different picture. The dependent measure was number of correct selections out of nine.

**Dot matrix oddity.** On each of seven pages, there were three 7.6 X 7.6 cm matrices. Each matrix was divided into nine squares. Two of the matrices had an equal number of dots placed in exactly the same position while on the different matrix, the dots were in the same positions but there was either one more or one less dot. Children were asked to point to the different design, the one not like the others. The dependent measure was number of correct selections out of seven.

**Sequencing.** On each of four pages, there was a sequence of either colored dots or animals running across the upper half of the page ending in a missing item. Children were asked to point to the one of three colored dots or animals, located on the bottom of the page that was missing from the top. The dependent measure was number of correct selections out of four.
**Picture rhyming.** Children were asked to point to the pair of pictures whose names rhymed, sounded alike. One pair appeared above a horizontal black line and the other below it. There was one training page and two test pages. The dependent measure was number correct out of two. Regardless of the language used to administer the instructions, the names rhymed in English, and were labeled in English.

**Unstructured semantic information.** Children were first shown a collage of 15 cats and then a grouping of fruit. They were asked to tell all they knew about cats and fruit. One point was awarded for each description that was valid for all, or most exemplars. The dependent measures were total number of points awarded to cats and to fruit.

**Structured information.** Pictures of four different people, who varied in race/ethnicity, age and gender, were displayed on a single page. Children were asked to tell how people differed from one another. One point was awarded for each valid difference. The dependent measure was total number of points awarded.

**Testing Procedure**

All children were tested individually, for a single session, by one of six female examiners. Before starting the testing session, examiners pointed out a selection of award certificates and stickers that the children were told they could earn by playing the game. Regardless of their performance, all children were given an award certificate and stickers. Presentation time was typically about 15 to 20 minutes. During presentation of the test, children were periodically praised for their good performance.

**Results**

Frequency distributions showing number of children in regular education and in special education who achieved each score were computed. At each cut score, all children with screening scores below the cut were said to be at risk for a special education placement, while those with scores at, or above the cut were predicted to be in regular education in third grade. Classification accuracy was determined by comparing the predicted to the actual educational placement of each child in third grade. The cut score selected was the one associated with the best combination of sensitivity and specificity (Yerushalmy, 1947). In this study, sensitivity was the percentage of all children with an EMH or SLD classification in third grade whose scores fell below the cut and specificity was defined as the percentage of all children in regular education in third grade whose scores were at, or above the cut.

**Classification Accuracy Using the Total Screening Score**

Since there were two measures to be evaluated for the picture pointing task, one frequency distribution used the quality score to represent the contribution of this task to the total score and the second distribution used the penalty measure. The best combination of sensitivity (82%) and specificity (79%) was achieved at a cut score of 60, using the penalty measure to represent the contribution of the picture pointing task to the total screening score. These values became the base against which estimates of sensitivity and specificity using subsets of tasks were compared.

**Classification Accuracy Using Subsets**

First, the best combination of sensitivity and specificity achievable for each individual task was determined. Since there were two measures for the picture pointing, unstructured semantic information and word meaning tasks, 12 measures were examined in this initial evaluation. The best combination of sensitivity (82%) and specificity (68%) was associated with scores from the structured information task.

Each child’s score on the structured information task was then combined with his/her score on each of the other measures, which resulted in the computation of 11 new frequency distributions, each using a score summed over two tasks. The best two-task set was chosen and additional measures were added (i.e., best three-task combination, four-task combination, etc.) until no further improvement in either sensitivity and/or specificity was obtained. In this instance, the highest levels of sensitivity (91%) and specific-
Sensitivity (85%) were achieved using a five-task set consisting of the structured information, standard oddity, word meaning/banana, dot matrix oddity and unstructured semantic information/fruit tasks, with a cut score of 21. The distribution for this subset is shown in Table 1. Scores have been blocked in order to illustrate the findings in a single table. This combination improved on the accuracy levels obtained with the full battery in terms of both sensitivity (91% vs. 82%) and specificity (85% vs. 79%).

Other Psychometric Characteristics

As well as evaluating the sensitivity and specificity one can achieve with a screening test, it is also important to know, for example, the percentage of underreferrals associated with the test. This measure was defined as the percentage of children with an SLD or EMH classification whose screening scores were at, or above the cut score (see Table 1). Only one of the eleven children (9%) with an SLD or EMH classification earned a score that was not below the cut and was, therefore, incorrectly predicted to be in regular education in third grade. In addition, only 15% of the children in regular education in third grade had kindergarten screening scores below the cut. These students are called false positives because they were incorrectly predicted to be in special education in third grade. Both of these values represent good psychometric characteristics.

However, predicting three years into the future led to an 84% over referral rate. That is, of all kindergarten children with scores below the cut who were predicted to be at risk for a special education placement, 84% were not in special education in third grade. Only 16% of children with scores below the cut actually did have an EMH or SLD label by third grade. This value represents the positive predictive accuracy of the set (Feinstein, 1976) and, on the face of it, is too high.

Over Referral and Poor Achievement

Percentile achievement scores on the reading comprehension, math computation and math application subtests of the Stanford Achievement Test (SAT) were available for most of the children. In the preschool predictive study (Scott & Delgado, 2003), many children in regular education who had preschool screening scores below the cut (false positives), ac-

---

**TABLE 1**

Number of Children in Regular Education and Special Education Achieving Each Score on the Five-Task Subset That Included the Structured Information, Standard Oddity, Word Meaning/Banana, Dot Matrix Oddity, and Unstructured Semantic Information/fruit Tasks

<table>
<thead>
<tr>
<th>Score</th>
<th>Regular Education</th>
<th>Special Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>8–10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>11–13</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14–16</td>
<td>11 (n = 52)</td>
<td>1 (n = 10)</td>
</tr>
<tr>
<td>17–19</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Cut Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21–23</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>24–26</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>27–29</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td>30–32</td>
<td>39 (n = 293)</td>
<td>0 (n = 1)</td>
</tr>
<tr>
<td>33–35</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>36–38</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>40–42</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>44–49</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>345</td>
<td>11</td>
</tr>
</tbody>
</table>

Sensitivity = 10 of 11 or 92%  
Specificity = 293 of 345 or 85%
tually did demonstrate learning difficulties in third grade. They earned very low achievement scores on one or more of the three achievement tests available in the data set. For the purpose of examining this aspect of academic performance, we defined poor achievement as having one or more achievement scores at, or below the 15th percentile. Children excluded were those with no SAT data and those with one or two missing scores whose recorded score(s) was greater than the 15th percentile. For example a child with one score at the 25th percentile would be excluded because we did not know what scores would have been achieved on the other two subtests; one or more might have been at or below the 15th percentile. However, if a single score was at or below the 15th percentile, even if one or more subtest scores were missing, then that child was classified as a poor achiever. Only children who could be classified as either a poor achiever or not were included in analyses of achievement data.

The same definition of poor achievement was used in this study. Of the 345 children in regular education, 340 could be classified as poor achievers or not. Overall, 79 or 23% were poor achievers in third grade. Of more importance to an evaluation of the over referral problem, 49 of the 52 children in regular education with screening scores of less than 21 could be classified. Of these 49, 28 or 57% were poor achievers in third grade. This represented approximately 35% of all children who were poor achievers in third grade.

If we now look at the predictive accuracy associated with selecting children at risk for learning problems in third grade, where learning problems are defined as either showing poor achievement or having an EMH or SLD label, then the positive predictive accuracy associated with the five-task set over three years is 64%, not 16%. This numerator for this computation was 38 which was the sum of the number of children with an EMH or SLD label (n = 10) plus the number of regular education children with scores below the cut who were poor achievers (n = 28). The denominator was 59, which was the sum of the total number of children in regular education with scores below the cut who could be classified (n = 49) plus the number of children with an EMH or SLD label with a score of less than 21 (n = 10). The percentage just drops to 61 if all children in regular education with scores below the cut are used, rather than only those who could be classified as a poor achiever or not.

Variables Impacting Kindergarten Screening Performance

In all of the following analyses, children’s screening scores on the five-task subset were used because this was the subset that maximized predictive accuracy. Unless otherwise indicated, the entire predictive sample of 356 were included because in kindergarten, when the children actually were administered the screening battery, none was in special education. However, all analyses of achievement performance included only the normally achieving group from the predictive sample because with this measure we were looking for children with poor achievement who were in regular, not special education.

Gender. Although males (M = 24.8; SD = 5.3) had significantly lower screening scores, F(1, 354) = 6.36, p < .02, than females (M = 26.2; SD = 5.2), the higher percentage of male (20.5%) than female (13%) children with scores below the cut was not significant, X^2 (1, N = 356) = 3.02, p > .05, nor was the difference between males (64%) and females (44%) in terms of the percentage of children with scores below the cut who were poor achievers in third grade, X^2 (1, N = 49) = 1.74, p > .05.

Testing language. We first compared the Hispanic/English sample (n = 101) with the Hispanic/Spanish sample (n = 19). The Hispanic/Spanish group (M = 20.9, SD = 5.8) had significantly poorer screening scores, F(1, 118) = 12.58, p < .01, than the Hispanic/English group (M = 25.5, SD = 5.1). There was also a significantly higher proportion of children from the Hispanic/Spanish group (37%) than from the Hispanic/English group (17%) who earned scores of less than 21, X^2 (1, N = 120) = 4.00, p < .05. These differences favoring Hispanic children tested in English, however, did not impact actual placement as all five Hispanic children in special education in third grade were tested in English.

Both the results pertaining to the screening
score measure, \(F(1, 123) = 12.35, p < .01\), and those reflecting the proportion of children with scores below the cut, \(X^2 (1, N = 125) = 5.03, p < .02\), were replicated, when the Hispanic/Spanish group was combined with the Hispanic/both group (\(n = 24\)) to form the Hispanic/Spanish-both group (\(M = 21.4, SD = 5.6\)).

Testing language did, however, impact achievement status. The combined Hispanic/Spanish-both group was used (\(n = 24\)) to increase the sample size for this categorical test. Considering only those children with scores below the cut, 23% of those tested in English were poor achievers compared to 78% of those children tested in Spanish or both languages. The difference in proportions was significant, \(X^2 (1, N = 22) = 6.42, p < .03\).

Race/ethnicity. Children in the White sample (\(M = 27.7, SD = 3.8\)) earned significantly higher screening scores, \(F(2, 353) = 10.74, p < .001\), than children in the Hispanic (\(M = 24.7, SD = 5.5\)) and Black (\(M = 24.6, SD = 5.5\)) samples, both \(p\) values < .001, while the two minority groups did not differ significantly from each other, \(p = .86\).

Children in the Black and Hispanic samples were also nearly six times more likely to earn a score below the cut, 23% and 21% respectively, than were children from the White sample (4%). For the race/ethnicity variable, the differential risk for a special education placement was validated. Of the 11 children classified as either SLD or EMH, 9% were White, but 45.5% were Black and 45.5% were Hispanic.

With respect to poor achievement, two children from the White sample had scores below the cut and one of these was a poor achiever in third grade. Such a small sample is unlikely to produce a reliable estimate. Consequently, a comparison across the three groups was not feasible. However, 68% of Black children and 45% of Hispanic children, with scores less than 21 were poor achievers in third grade.

The impact of testing language on race/ethnicity. Did the Hispanic sample as a whole perform at a lower level than the White sample or was it only those Hispanic children who were tested in Spanish or both Spanish and English? To answer this question, the Hispanic sample was divided into two groups: Hispanic/English (\(n = 101\)) and Hispanic/Spanish (\(n = 19\)). The other two racial/ethnic groups were also delineated in terms of their testing language: White/English (\(n = 85\)) and Black/English (\(n = 146\)).

An analysis of variance (ANOVA) of groups was significant, \(F(3, 347) = 11.71, p < .001\). Post hoc tests showed that while the screening scores of the White/English group were still significantly higher than the scores of all three minority groups, all \(p\) values < .01, both the Hispanic/English (\(M = 25.5; SD = 5.1\)) and Black/English (\(M = 24.6; SD = 5.5\)) groups had significantly higher screening scores than the Hispanic/Spanish group (\(M = 20.9; SD = 5.8\)), both \(p s < .01\). The same results were observed when children tested in both English and Spanish were added to the Hispanic/Spanish group.

A similar pattern of mean differences was observed when the percentage of children with scores below the cut was examined. The White/English sample had the fewest (4%), the Hispanic/English and Black/English had percentages in the middle, 17% and 23% respectively, and 37% of the Hispanic/Spanish group earned scores below the cut. There was a similar percentage of children with scores below the cut (38%) for the Hispanic/Spanish-both group. However, remember that of the five Hispanic students in special education, all were tested in English.

A slightly different pattern of group differences was observed when the dependent measure was the percentage of children in regular education, with scores below the cut, who were poor achievers in third grade. Only two children from the White/English sample had scores below the cut, so no reasonable estimate can be made of the percentage of poor achievers for this group. However, only 23% of children from the Hispanic/English group with scores below the cut were poor achievers while 68% of children in the Black/English group and 78% of children in the Hispanic/Spanish-both group with scores below the cut were poor achievers in third grade.

SES. SES of each child was estimated using the percentage of children eligible for free or reduced lunch in his/her school. A child was eligible for reduced lunch if annual household income was at, or less than, 185% of the poverty level, which is defined annually by the federal government. A child was eligible for
free lunch if annual household income was 130% of the poverty level. Since eligibility is in terms of household income level, the percentage of children in each school who are eligible for these services is a good estimate of the SES of the families in that catchment area.

Schools were divided into three SES groups; low, mid-level and high. Schools were designated as serving a low SES population if from 70.6% to 99% of students in that school were eligible for free or reduced lunch. Schools were placed in the mid-level SES group if from 30.6% to 70.5% of their population were eligible for free or reduced lunch. Schools with less than 30.6% of their population eligible for free or reduced lunch were designated high SES. There were 15, 15 and 3 schools in the low, mid-level and high SES conditions, respectively.

A one-way ANOVA of SES group was computed with scores on the five-task subset as the dependent measure. Mean screening score increased as SES increased from low (24.2; SD = 5.6), through mid-level (25.9; SD = 5.4) to high (26.7; SD = 3.8). These differences were significant, $F(2,353) = 5.89$, $p < .01$. Post hoc tests showed that both the high and mid-level SES condition were associated with higher screening scores than the low SES condition, both $p$ values < .01, but were not significantly different from each other, $p = .29$.

The percentage of children with screening scores below the cut increased as SES level decreased. Only five percent of children in the high SES group, but 14% of children in the mid-level group and 21% of children attending low SES schools earned such low kindergarten screening scores. Similarly, 9% of the children in special education were in a high SES school in kindergarten compared to 45.5% who were in mid-level SES schools and 45.5% who were in low SES schools.

A similar relation was present in terms of the percentage of children with kindergarten screening scores below the cut who were poor achievers in third grade. Again, only three children from the high SES condition had scores below the cut and one of these was a poor achiever in third grade; a small sample to estimate from. However, 57% (12 of 21) of children with scores below the cut in the mid-level SES condition, and 60% (15 of 25) of children with scores below the cut in the low SES condition were poor achievers in third grade.

Impact of SES on race/ethnicity. Since there were so few Hispanic children tested in Spanish ($n = 19$) or both languages ($n = 5$), these two groups were combined when the impact of SES on screening performance was examined. Number of children in the White/English, Hispanic/English, Black/English and Hispanic/Spanish-both groups who were in each of the three SES conditions was examined. There were eight or fewer students in five cells of the twelve-cell matrix. This precluded using an ANOVA. However several relationships between SES and other variables were observed.

With respect to race, the White/English sample had the highest percentage (40%) of students in high SES schools while not one student in the Black/English group attended a high SES school. In contrast, only 5% of the White sample but 77% of the Black/English sample were in low SES schools.

With respect to ethnicity and testing language, more children from the Hispanic/English (24%) than from the Hispanic/Spanish-both sample (12.5%) were attending high SES schools, while fewer children in the Hispanic/English sample (10%) than in the Hispanic/Spanish-both sample (29%) were enrolled in low SES schools.

Discussion

The major intent of this study was to evaluate a kindergarten screening battery, and selected items from that battery, in terms of the accuracy achieved when kindergarten screening scores were used to predict third grade educational status.

Children’s total scores on the kindergarten screening battery were associated with a sensitivity of 82% and a specificity of 79%. However, consistent with previous data (e.g. Scott, Deuel, et al., 1998, Stebbins & McIntosh, 1996), a higher level of both sensitivity (91%) and specificity (85%) was achieved with a five-task subset consisting of the structured information, standard oddity, word meaning/banana, dot matrix oddity and unstructured semantic information/fruit tasks. The inclusion of three generating tasks in this five-task subset is consistent with the position adv
cated by Simmons and Kameenui (1990) that performance differences between intact and mildly impaired groups will be greater when those groups are compared using tasks that a) require the participants to produce responses and b) have little external support. Generating tasks fit this description.

Both the sensitivity and specificity associated with this five-task set are above the minimum 80% levels recommended by Meisels (cited in Lichtenstein & Ireton, 1991). They are also exceptionally high when considering the influence on school performance of potentially huge variations in the home and school environments the children were likely to have experienced during the three years between the kindergarten assessment and third grade. Varied experiences can be expected to differentially affect performance and contribute to the difficulty of achieving accurate long term prediction (Lichtenstein & Ireton). Indeed, these authors suggested that with preschool screening, perhaps one must settle for a “reasonable correspondence with a concurrent developmental assessment” (p.503).

Results of this study show, however, that early and accurate identification of kindergarten children who will have mild learning problems later in school is possible.

Although nearly all children with a special education placement of either SLD or EMH were accurately identified, there was a very high level of over referral. A high rate of over referral is often found with developmental screening measures (Gredler, 1997; Lichtenstein & Ireton, 1984) and may be a necessary consequence of trying to identify a low prevalence group. Another factor bearing on the actual impact of the high percentage of over referrals to the utility of the screening set is the finding that of the children in regular education with scores below the cut whose achievement status could be ascertained, 57% were poor achievers. In short, a high percentage of false positives using the placement-in-special education criterion were actually evidencing poor achievement, another type of learning problem. These results replicate those reported in Scott & Delgado (2003) for a preschool predictive sample where there was also a high percentage of false positives who were poor achievers.

If one considers all those children with scores below the cut of 21 who were either in special education or were poor achievers, then the positive predictive accuracy associated with the identification of children at risk for either poor achievement or a special education placement is 64%, a value much higher than the 16% associated with risk for special education placement alone. Indeed, if the most critical aspect of a screening test is its ability to identify those children who will have problems in school (e.g., Limbos & Geva, 2001; Rafoth, 1997), the five-task subset identified in this study can be considered to be a successful and effective screening device.

Since 36% of kindergarten children with scores below the cut were neither in special education nor were poor achievers in third grade, one is hesitant to refer all children with scores below the cut for a costly psychoeducational assessment. Rather, results of the children’s performance on this very brief screen could be used to identify a subsample of children who are at risk for school problems, a group who should be carefully monitored for possible learning problems. Then, at the earliest sign of any academic problems by anyone from this risk group, the combination of risk status and observed academic difficulties should immediately indicate the provision of a prereferral intervention to the target child, rather than waiting for further educational difficulties to manifest themselves. The provision and careful monitoring of such interventions has been shown to result in academic improvement (Quiroga, Lemos-Britton, Mostafapour, Abbott, & Berninger, 2002), which should minimize the extent and/or depth of the problems that could occur later.

Monitoring, rather than placement, would incur no additional costs, nor would it lead to the improper labeling of any child. Rather it would simply serve as an early warning signal of potential, not certain, academic problems that might eventually require remedial instruction or for some, a special education placement.

What factors increased the risk for achieving a score of less than 21? A relation between poor cognitive performance and minority status has been previously demonstrated in similar screening studies (e.g., Scott, Fletcher, Jean-Francois, et al., 1998; Scott et al., 2003) and minority status had a major negative im-
pact on performance again in this study. Hispanic and Black children were more than five times more likely to earn a score below the cut and five times more likely to be in special education compared to White children. It should be noted, however, that both SLD and EMH are low prevalence conditions so that overall, only 3.4% of the Black sample and 4% of the Hispanic sample actually ended up in special education.

A high percentage of those Hispanic and Black children with scores below the cut were also poor achievers. This outcome was more prevalent overall than was a placement in special education, with 18% of the total Hispanic sample and 36% of the total Black sample meeting the criterion for poor achievement in third grade.

Negative impact of minority status was mitigated for Hispanic children on some outcome measures if they were tested in English rather than Spanish or both languages. Children from the Hispanic/English group were less likely to have a kindergarten screening score below the cut and if they did have a score below the cut, were much less likely to be a poor achiever in third grade. Indeed, the Hispanic/English sample looked much more like the nonminority, White group, than did either of the other two minority groups.

Therefore, any Hispanic child who chooses (needs) to be tested in Spanish or both languages and who achieves a score below the cut should be rapidly responded to at the first sign of any achievement problems. A prereferral intervention geared to foster relevant cognitive processing skills and to improve English language performance and comprehension might be a cost effective approach to use with this particular group.

The increased risk for obtaining a screening score below the cut observed for the Hispanic/Spanish and Hispanic/Spanish-both groups was not related to educational status in third grade. Of the five Hispanic children in special education, none was tested in Spanish or both languages. All were tested in English. Absence of a negative impact of testing in other than English on this measure is probably related to the fact that the total number of children who were not tested in English was only 24 compared to the 101 who were tested in English. Given the low prevalence of a special education placement in the two categories examined, 4.5% over grades one through six, it is not surprising that all five Hispanic children who were in special education in third grade came from the larger sample.

The relation between SES and screening performance proved to be a major confound of the race/ethnicity effect. Nearly half the children in special education were Black and of those Black children with scores below the cut, 68% were poor achievers. However, it is likely that the poor performance of this group is related not to race per se, but to the fact that 77% of the Black sample lived in low SES catchment areas. Not one Black child in this sample was attending a high SES school. The confounding of minority status and low SES is consistent with other reported findings (Halle et al., 1997) as is the relation between low SES and poor achievement (Hill, 2001). All the characteristics that project to an expectation of poor academic performance are most likely to correlate with poverty, e.g., lower education in the parents, less verbal interactions between parents and children, lack of or little written material in the home. It would seem reasonable to conclude that the poorer performance of this particular minority group was drastically affected by factors associated with low SES.

Similarly, the superior performance of the Hispanic/English group compared to the Hispanic/Spanish or Hispanic/Spanish-both groups is partly confounded by SES. The percentage of the Hispanic/English group who lived in a high SES catchment area was twice as large as the percentage of the Hispanic/Spanish-both group. In contrast, the percentage of the Hispanic/Spanish-both group who were attending low SES schools was nearly three times as large as the percentage of children from the Hispanic/English sample who lived in a low SES catchment area.

Overall, it must be noted that on nearly all measures the Hispanic/Spanish and Hispanic/Spanish-both groups also performed more poorly than the Black group.

These data can be contrasted with a much poorer educational outcome for the Hispanic/Spanish and Hispanic/Spanish-both groups seen in the preschool predictive study (Scott & Delgado, 2003) when more than 84% of children in both groups were attending low SES schools. Clearly SES is a confounding vari-
able that must be taken into account when comparing racial/ethnic groups.

Since this screening set was selected based on the specific performance of this particular group of children, a cross validation study is necessary as are a number of other evaluative studies, such as an assessment of the test-retest reliability of the screening set. However, these data do show that it is possible to identify nearly all kindergarten children with only mild learning problems who will require special education services at a later date and a little more than a third of the children who will have achievement problems later in school.

Early identification of kindergarten children at risk for academic problems regardless of severity is critical to successful remediation efforts. This brief screening set, which is easy to administer and takes little time, will allow one to target a subset of children, the majority of whom are at greatest risk for a poor educational outcome. Monitoring the actual performance of this limited subset of children in order to detect any early signs of learning difficulties would be an inexpensive method of attacking a major educational problem.

References


Early identification: An analysis of the research. Learning Disability Quarterly, 11, 176–188.

Received: 28 April 2004
Initial Acceptance: 29 June 2004
Final Acceptance: 20 December 2004