Evaluation of the young child and his or her family is one of the most basic aspects of occupational therapy services in pediatrics and includes both screening and diagnostic instruments (Dunn, 2000a). The goal of screening is to identify children who may be candidates for in-depth testing of whether they could benefit from preventive or remedial intervention programs (Carran & Scott, 1992; Dunn, 2000a; Gredler, 1997; May & Kundert, 1997; Miller, 1988a, 1988c). As such, screening is an example of occupational therapy service delivery within the context of the current health care paradigm that emphasizes preventive services (Dunn, 2000a), maintenance and enhancement of the individual's functional status (Baum, 2000), and consideration of cost–benefit ratios in care (Wenner, 1995).

Screening is especially relevant for the families of young children because research supports that early intervention enhances long-term functioning in all occupational areas (Gallagher, 1993; Henderson & Hall, 1982; McLaughlin & Wehman, 1992). Effective screening can facilitate family-centered therapy (Dunn, 2000a) in that parents’ knowledge of their child's condition leads to parent involvement in evaluation and therapy and, as a result, to enhanced functional outcomes (Calnan et al., 1994; King, King, & Rosenbaum, 1996; Rosenbaum, King, Law, King, & Evans, 1998).

Evidence-based practice demands the selection of the current best tools (Tickle-Degnen, 1999). Pragmatic characteristics, such as ease of administration and appropriate content domains, are crucial (Gredler, 1997; Meisels & Wasik, 1990). Another feature of an effective screening tool is that it should have at least one cut-point below which children are considered “at risk” and in
need of referral for further assessment (Crossland, 1994).

Another critical issue in selecting an evidenced-based screening tool is the determination of appropriate levels of sensitivity and specificity. These data are essential to evaluating the accuracy with which the test detects the target population. Sensitivity indicates the extent to which the test detects dysfunction when it is actually present; specificity indicates the extent to which the test accurately rules out dysfunction. False negatives result in a lack of special services to children who could benefit from the services; false positives result in identifying children who are actually typically developing as at risk. Both types of mislabeling can have serious consequences.

A primary concern in the selection of a screening instrument is predictive validity, which refers to the tool’s ability to predict accurately persons who will develop future problems versus those who will not (Barnes, 1982; Lichtenstein & Ireton, 1991). Predictive validity requires that longitudinal analysis be accomplished through the use of follow-up studies (Carran & Scott, 1992). Follow-up studies that compare screening test scores with later target performance abilities of similar subjects evaluate the screening prediction relative to the later measured outcome (Miller, Lemerand, & Schouten, 1990). A screening instrument that demonstrates a high level of accuracy after a long interval is the most useful (Barnes, 1982).

The literature on preschool screening is replete with calls for longitudinal research; however, the predictive validity of only a few preschool screening tests is available (Badian, 1988; Gredler, 1997). Simner (1996) found that kindergarten children with lower scores on the Caregiver’s School Readiness Inventory (Simner, 1989), a nontraditional preschool screening for school failure, were likely to exhibit poor academic performance at the end of second grade. Kelly and Peverly (1992) demonstrated that the Kindergarten Screening Battery (Belkin & Sugar, 1985) was an effective predictor of academic achievement in reading, mathematics, and listening comprehension in first and second graders. The Brigance K and 1 Screen (Brigance, 1982) and the Merrill Language Screening Test (Mumm, Secord, & Dykstra, 1980) were predictive of kindergarten teachers’ recommendations for referral to remedial programs and for retention (Wennen, 1995). The Early Screening Inventory (Meisels, Henderson, Liaw, Browning, & Have, 1993) and the Early Prevention of School Failure (Roth, McCaul, & Barnes, 1993) also have shown predictive utility, particularly in fine motor and auditory modalities. The Revised Denver Developmental Screening Test (Frankenburg, Fandel, Sciarillo, & Burgess, 1981) was found to be an efficient prognostic tool of formal assessment results for 3- to 4-year-old preschool children at risk for language impairments over a 6-month test interval (Feene & Bernthal, 1996).

The Miller Assessment for Preschoolers (MAP; Miller 1988c) is a widely used preschool screening test. The MAP taps into sensory, motor, perceptual, cognitive, and verbal performance components in children and contains activities that are easily administered in an appealing game format.

Several longitudinal studies have been performed in the United States with the purpose of establishing the predictive validity of the MAP. Cohn (1986) performed a follow-up study after an interval of 2 years from screen testing with the MAP. In Cohn’s study, the MAP’s prediction of at risk versus not at risk was evaluated in relation to teachers’ ratings of overall academic performance and assignment to special services in first grade. Results indicated that the means of the MAP scores distinguished between groups of at-risk and not-at-risk children in their first school year in accordance with teachers’ ratings. In another study, Lemerand (1988) found that the classification analysis of the MAP’s 25% cut-point, after a 1-year interval, indicates sound sensitivity and specificity rates (2.070) in predicting children who later had difficulty in kindergarten.

Miller (1988a, 1988b) conducted a follow-up study of the MAP standardization sample and examined the results of intelligence and achievement tests conducted on the participants 4 years after screening (Miller, 1988b). She found that the highest correlation was between the MAP total score and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991) criteria (r = .45–.50) (Miller, 1988c). Strong relationships also were found between the MAP and achievement tests, such as the Woodcock measures (Woodcock & Johnson, 1977) (r = .35–.38). These studies demonstrate that the MAP, when given to preschoolers, adequately predicted intelligence and academic performance 4 years later (Miller, Lemerand, & Cohn, 1987).

The purpose of the present study was to cross-validate the previous findings that the MAP is a screening tool capable of predicting performance of school-age children over a substantial period (7 years). Therefore, the assessments selected for the follow-up testing were designed to tap into a wide range of performance components that spanned the areas of motor, perception, cognition, and functional academic performance, such as reading and writing. Because the normative data for the MAP were obtained in 1980, this study attempts to provide evidence to update and substantiate previous claims of predictive validity.

Moreover, comparing screening results with relevant aspects of future functional status gives us insight into the relationship between performance components and functional outcomes, a concept about which many have theo-
rized (Case-Smith, 2000; Coster, 1998) yet for which almost no empirical data can be found. It is noteworthy that this study was conducted on Israeli children and used the Hebrew version of the MAP, which was standardized on a sample of Israeli preschool children (Schneider, Parush, Katz, & Miller, 1995). The Israeli standardization study found no significant differences with respect to MAP total scores between the norms established by Miller on American preschoolers and Israeli preschoolers. Therefore, this follow-up study that examines functional outcomes and occupational performance of school children from another culture affords us additional insight about the functioning of school-age children irrespective of the cultural context.

Method

Design

Establishing predictive validity of a screening tool is a process that is tested over time. This research is a follow-up study examining academic outcome and performance components of school children who had been previously classified (5–7 years before) through the MAP as being at risk or not at risk for pre-academic problems.

Sample

A convenience sample of 30 children (14 girls, 16 boys) in grades 4 through 7 from two cities in central Israel participated in the present study. They were selected from a total of 128 MAP test protocols that had been administered to all the children registered in four Mother and Child Health Care centers. The original screening was administered as part of an early screening project carried out between the years 1989 and 1992, with follow-up occurring 5 to 7 years later.

The 30 children were divided into two groups according to the risk criteria devised by MAP total scores. The children in the not-at-risk group (n = 15; 9 girls, 6 boys) were within normal limits on the MAP on the original screening (i.e., scores between 25% and 99%). The children in the at-risk group (n = 15; 5 girls, 10 boys) had MAP scores between 1% and 25%. The mean age for the sample was 3 years, 10 months, at the time of the original screening. At follow-up, the mean age was 11 years, 11 months, for the at-risk group and 11 years, 4 months, for the not-at-risk group (a nonsignificant group difference). All children in both groups were living with their two-parent families (biological or legal) at the time the research took place.

Socioeconomic status (SES) was established through Hartman’s (1975) formula and grading, which takes into consideration both parents’ educational background and the father’s present occupation. No significant differences between groups were found in these demographic variables (see Table 1).

Instrument

The MAP score (Miller, 1988c) was the independent variable. The MAP is an individually administered test that identifies children between the ages of 2 years, 9 months, and 5 years, 8 months, who may be at risk for mild to moderate pre-academic problems. It examines performance components that may affect school function and includes 27 subtests in 5 domains: neurological foundations, motor coordination, language, nonverbal cognition, and complex tasks (combined domains). The total MAP score is expressed in percentiles, and the cut-points are 0% to 5% (Red; likely problem, refer for evaluation), 6% to 25% (Yellow; possible problem, watch carefully and use clinical judgment about the need to refer for evaluation), and 26% to 99% (Green; unlikely to have problems, do not refer for evaluation).

The MAP has excellent internal reliability (r = .79–.82) and interrater reliability (r = .98) (Daniels & Bressler, 1990; Miller, 1988c). Test–retest reliability for total score is r = .81 (Miller, 1988c). Content validity for the MAP is supported in the literature as MAP total score correlates significantly with the WISC-R IQ scale (r = .50–.45) and with the Woodcock-Johnson Math, Reading and Language subtests (r = .38–.35) (Miller, 1988c). In the present study, the translated version of the MAP was used for which test–retest reliability for the total score is r = .74, and interrater reliability is r = .91 (Schneider et al., 1995). The follow-up dependent measures were as follows.

Motor performance components. The Bruininks-Oseretsky Test of Motor Proficiency (BOTMP), Short

Table 1. Demographic Characteristics of Children in the At-Risk and Not-At-Risk Groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>At Risk</th>
<th>Not Risk</th>
<th>Z*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>School grade</td>
<td>5.47 ± 0.99</td>
<td>5.33 ± 0.98</td>
<td>-433</td>
<td>NS</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>2.67 ± 0.90</td>
<td>2.67 ± 1.05</td>
<td>-116</td>
<td>NS</td>
</tr>
<tr>
<td>Siblings with learning problems</td>
<td>0.33 ± 0.49</td>
<td>0.33 ± 0.49</td>
<td>0</td>
<td>NS</td>
</tr>
<tr>
<td>Birth order</td>
<td>2.07 ± 0.96</td>
<td>2.00 ± 0.93</td>
<td>-131</td>
<td>NS</td>
</tr>
<tr>
<td>Socioeconomic status</td>
<td>2.95 ± 0.92</td>
<td>3.08 ± 1.16</td>
<td>-438</td>
<td>NS</td>
</tr>
</tbody>
</table>

*Mann Whitney U test.
Form (Bruininks, 1978) is an individually administered test that assesses the gross and fine motor performance components of children from 4 1/2 to 14 1/2 years of age. The Short Form consists of 14 items from the complete battery. It contains a brief survey of general motor proficiency: running speed and agility, balance, bilateral coordination, strength, upper-limb coordination, response speed, visual-motor control, and upper-limb speed and dexterity. Interrater reliability ranges from $r = .80$ to .97, and test–retest reliability ranges from $r = .86$ to .97 (Bruininks, 1978).

**Visual-motor performance components.** The Developmental Test of Visual-Motor Integration (VMI; Beery, 1989) is a developmental sequence of 24 geometric forms to be copied with paper and pencil. The age range includes preschool children through adults. This scale correlates significantly with academic achievement (correlations ~ .50). Interrater reliability is $r = .93$, and test–retest reliability ranges from $r = .75$ to .92 (Beery, 1989).

**Cognitive performance components.** The Pictorial Sequence subtest of the Loewenstein Occupational Therapy Cognitive Assessment (LOTCA; Itzkovich, Elazar & Averbuch, 1993) comprises two series of pictorial sequences that assess the sequential thinking performance of children. The children are asked to arrange picture cards according to what they consider to be a logical sequence and to describe the resulting story. Each subtest is scored on a scale from 1 (unable to perform) to 4 (good performance). Interrater reliability ranges from $r = .82$ to .97, and the alpha coefficient for internal consistency is .85. Construct validity has been demonstrated (Averbuch & Katz, 1991).

The Visual Aural Digit Span Test (VADS; Koppitz, 1977) measures both cross-modal sensory integration and short-term memory performance components. This test detects learning problems in children from 5 years, 6 months to 12 years of age. It consists of four subtests: Aural-Oral, Visual-Oral, Aural-Written, and Visual-Written. Reliability for subtest scores ranges from $r = .72$ to .90. The VADS score for kindergarten children has concurrent validity for second through fifth grade children's reading achievement scores (Koppitz, 1977).

The Conners Abbreviated Symptom Questionnaire (ASQ; Goyette, Conners, & Ulrich, 1978), a parent rating scale, was completed by the parents in order to detect attention deficits and hyperactivity. This 10-item index reflects the most frequently endorsed items in Conners’s more extensive Teacher and Parent Rating Scales (Conners, 1985). Scaling for each item ranges from 0 to 3, and total scores above 15 are considered to suggest hyperactive behavior. The appropriate ages for administration of this questionnaire range from 3 years to 17 years. Interrater reliability is not reported; however, interparent agreement is $r = .55$ to .71, and parent–teacher agreement is $r = .49$ (Goyette et al., 1978).

**Academic performance area.** The Hebrew Handwriting Evaluation (HHE; Erez & Parush, 1999) was administered to assess handwriting proficiency, handwriting speed, self-comprehension, and subjective legibility and spelling in copying and dictation conditions. A total score is calculated for Handwriting–Copying and Handwriting–Dictation by summing the scores of the different relevant subtests. Good interrater ($r = .75$–.79) and internal ($r = .81$) reliability correlations have been established for this scale, and construct validity has been demonstrated (Devash, Levi, Traub, & Shapiro, 1995).

Reading ability was tested by the diagnostic–didactic tests developed and used at Nizan, the Israeli Center for Learning Disabled Children. The tests measure a child's level of proficiency in silent reading and in recitation (Kidron, 1990) in the Hebrew language. A total score was calculated for reading proficiency by summing the scores of the different reading subtests (silent, recitation). Although these tests are widely used in Israel, no reliability or validity data have been reported.

**Child placement and intervention information.** A parent questionnaire requiring yes and no responses was used to gather data about the following dependent variables: (a) rate of special education attendance currently or during kindergarten, (b) rate of kindergarten retention, and (c) number of special service interventions (speech therapy, occupational therapy, psychology, etc.) the children received.

**Procedure**

The original MAP screening was administered by two qualified occupational therapists who were trained by an expert examiner in a special workshop for MAP administration. The current sample for the follow-up testing was selected in three stages: (a) examination of MAP protocols, (b) tracing the potential participants, and (c) procuring agreement (informed consent). MAP scores were obtained by retrospective analysis of test protocols from the centers’ archives. Children who had incomplete MAP protocols were excluded. One hundred twenty-eight children met entrance criteria of which 69 families were located. Of the 69 families, 26 had children who were identified as at risk according to MAP criteria, and 15 agreed to participate in the study. From the not-at-risk children (43), a control sample of 15 was matched for age and SES to the at-risk children.

Each follow-up test session lasted approximately 2 hours. All testing was done by the second author at the child’s home during the afternoon hours. She also adminis-
tered the reading tests under the supervision of two didactic testing professionals. It is noteworthy that the second author did not participate in the original MAP screening, eliminating a possible source of experimenter bias.

Data Analysis

All data were expressed as raw scores because the standard scores and percentiles of the different measures were not standardized on Israeli children. It is noteworthy that raw scores are considered a more reliable indication of change over time (Wilson, Polatajko, Kaplan, & Faris, 1995); therefore, use of the raw score may facilitate further research with these children.

Depending on the level of the measurement scale, the variable Mann Whitney $U$ and $t$ tests were used to compare the demographic and test performance data of the two groups across time. To determine the extent to which our results might depend on our sample size, post hoc power estimates were calculated according to the observed sample mean difference, the sample standard deviations, and the sample sizes. Results are interpreted as the probability that this effect size will be detected; the higher the obtained number, the higher the probability of detection of a given effect.

Results

The MAP total score (in percentiles) ranged from 3 to 25 for the at-risk group and from 47 to 92 for the not-at-risk group. The mean total scores were 16.13% ($SD = 7.77$) for the at-risk group and 73.33% ($SD = 18.57$) for the not-at-risk group. Meaningful differences were seen between both groups with regard to child placement and intervention services. In the not-at-risk group, all the children attended regular kindergartens, and none were retained for a second year. Moreover, none of the children in this group required special education services, and only 4 received developmental special service interventions. In contrast, 4 children in the at-risk group were placed in special education kindergartens. Three of these children continued on in special education frameworks, with 2 in special education classes within regular schools and 1 attending a segregated special education school. This child had the lowest MAP score from within the follow-up research sample. In addition to the 4 children in special education, 10 children in the at-risk group received special service intervention in one developmental area (e.g., speech therapy), and 7 children were receiving two types of treatment simultaneously (e.g., occupational therapy and physical therapy).

Many of the outcomes for the tests administered in this study—BOTMP, VADS, VMI, Conners ASQ, LOTCA, and handwriting and reading scales—demonstrated significant group differences, with the not-at-risk group performing better on these measures than the at-risk group. Raw scores in the not-at-risk group indicated better performance than those of the at-risk group on all tests. To correct for the number of $t$ tests, we adjusted the alpha level to .01. Group differences were significant at the .01 level or less in 13 of the 26 subtests. Tables 2 and 3 compare the groups’ performances on all the follow-up tests. They indicate that significant differences exist, especially in cognitive and visual-motor performance components (i.e., LOTCA, VMI) and academic performance area (handwriting, three of five subtests in reading) (see Table 3).

To determine the extent to which these results might depend on the size of the sample, post hoc calculations of power estimates were done (Cohen, 1992). These estimates yielded the following results: BOTMP, .14; Conners ASQ, .30; VADS, .50; VMI, .87; LOTCA, .98; HHE (copy), .98; HHE (dictation), .87; and reading proficiency, .87. These data indicate that had a larger sample size been used, greater group differences would have resulted on the VMI, LOTCA, HHE (copy and dictation), and reading proficiency. In contrast, a larger sample size would not have

<table>
<thead>
<tr>
<th>Test</th>
<th>At-Risk Group</th>
<th>Not-At-Risk Group</th>
<th>$t$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOTMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw score</td>
<td>64.40</td>
<td>67.40</td>
<td>-1.004</td>
<td>28</td>
<td>.324</td>
</tr>
<tr>
<td>Standard score</td>
<td>51.13</td>
<td>58.20</td>
<td>-1.252</td>
<td>28</td>
<td>.221</td>
</tr>
<tr>
<td>VADS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aural-Oral</td>
<td>5.53</td>
<td>6.20</td>
<td>-1.967</td>
<td>28</td>
<td>.059</td>
</tr>
<tr>
<td>Visual-Oral</td>
<td>6.20</td>
<td>6.47</td>
<td>-0.907</td>
<td>28</td>
<td>.372</td>
</tr>
<tr>
<td>Aural-Written</td>
<td>5.13</td>
<td>5.93</td>
<td>-1.915</td>
<td>28</td>
<td>.066</td>
</tr>
<tr>
<td>Visual-Written</td>
<td>5.53</td>
<td>6.33</td>
<td>-2.806</td>
<td>28</td>
<td>.009</td>
</tr>
<tr>
<td>Total scores</td>
<td>22.40</td>
<td>24.93</td>
<td>-2.118</td>
<td>28</td>
<td>.043</td>
</tr>
<tr>
<td>VMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>31.07</td>
<td>38.87</td>
<td>-3.273</td>
<td>28</td>
<td>.003</td>
</tr>
</tbody>
</table>

Note. BOTMP = Bruininks-Oseretsky Test of Motor Proficiency; VADS = Visual Aural Digit Span Test; VMI = Developmental Test of Visual-Motor Integration.
changed the differences between the scores of the groups on the BOTMP, the Conners ASQ, and the VADS.

Sensitivity and specificity were calculated only for those tests that were found to be significantly different between groups. Table 4 displays the sensitivity and specificity values of the tests that were found to be significant in predicting children’s future performance (i.e., VMI, reading proficiency, HHE [copy and dictation]).

### Discussion

The findings from this study are consistent with those from the longitudinal studies undertaken in the United States to establish the predictive validity of the MAP with American children (Humphry & King-Thomas, 1993; Miller, 1988a). The longitudinal studies found consistent group differences between children at risk and not at risk in class retention, assignment to special services, and assignment to special education classes (Cohn, 1986; Lemerand, 1988; Miller, 1988a; Miller et al., 1987). As Miller (1988a) stated, “The MAP Total Score effectively differentiated problem and no-problem children” (p. 13).

In addition to cross-validating previous U.S. results, the current findings indicate that the not-at-risk group performed significantly better than the at-risk group in the perceptual, cognitive, and functional academic performance areas that were tested. Specifically, the not-at-risk group had higher scores than the at-risk group on the visual-motor integration, pictorial sequencing and narrative, the auditory and visual short-term memory, and the handwriting and reading proficiency tests.

No significant differences were detected in motor proficiency between the study groups on the BOTMP at follow-up. Motor proficiency, however, is reported to be an important indicator of future school success (Ellwein, Walsh, Eads, & Miller, 1991; Roussounis, Gaussen, & Stratton, 1987). We believe that methodological factors related to instrumentation might account for this finding. The BOTMP Short Form, used to avoid overtiring the children during testing sessions, has some limitations compared with the longer BOTMP. Because the short form contains fewer items, it might not be sensitive enough to detect a problem in a specific motor area. Additionally, the overall score of the BOTMP Short Form provides only one composite result without subtest results. Therefore, it can only indicate the existence of general problems. Further, because

### Table 3. Comparison of Scores—Conners ASQ, LOTCA, and Handwriting and Reading Proficiency

<table>
<thead>
<tr>
<th>Test</th>
<th>At-Risk Group</th>
<th>Not-At-Risk Group</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conners ASQ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total point score</td>
<td>10.81</td>
<td>7.27</td>
<td>−1.657</td>
<td>.098</td>
</tr>
<tr>
<td><strong>LOTCA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total point score</td>
<td>5.73</td>
<td>7.53</td>
<td>−3.487</td>
<td>.000</td>
</tr>
<tr>
<td><strong>HHE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>2.73</td>
<td>3.53</td>
<td>−3.102</td>
<td>.002</td>
</tr>
<tr>
<td>Self-comprehension</td>
<td>3.40</td>
<td>3.87</td>
<td>−2.318</td>
<td>.000</td>
</tr>
<tr>
<td>Subjective legibility</td>
<td>2.67</td>
<td>3.53</td>
<td>−3.510</td>
<td>.000</td>
</tr>
<tr>
<td>Spelling</td>
<td>2.80</td>
<td>3.87</td>
<td>−3.405</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Dictation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>2.67</td>
<td>3.47</td>
<td>−2.573</td>
<td>.100</td>
</tr>
<tr>
<td>Self-comprehension</td>
<td>3.13</td>
<td>3.73</td>
<td>−2.645</td>
<td>.008</td>
</tr>
<tr>
<td>Subjective legibility</td>
<td>2.47</td>
<td>3.33</td>
<td>−2.849</td>
<td>.004</td>
</tr>
<tr>
<td>Spelling</td>
<td>2.47</td>
<td>3.53</td>
<td>−3.518</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Reading proficiency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>3.27</td>
<td>3.73</td>
<td>−1.963</td>
<td>.050</td>
</tr>
<tr>
<td>Comprehension</td>
<td>2.80</td>
<td>3.67</td>
<td>−3.121</td>
<td>.002</td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.00</td>
<td>3.67</td>
<td>−2.408</td>
<td>.016</td>
</tr>
<tr>
<td>Silent reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>3.40</td>
<td>3.87</td>
<td>−1.998</td>
<td>.046</td>
</tr>
<tr>
<td>Comprehension</td>
<td>2.87</td>
<td>3.73</td>
<td>−3.084</td>
<td>.002</td>
</tr>
</tbody>
</table>

Note. Conners ASQ = Conners Abbreviated Symptom Questionnaire; LOTCA = Loewenstein Occupational Therapy Cognitive Assessment (the Pictorial Sequence subtest); HHE = Hebrew Handwriting Evaluation.

### Table 4. Sensitivity and Specificity of the Miller Assessment for Preschoolers

<table>
<thead>
<tr>
<th>Criterion Test</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMI</td>
<td>85.71</td>
<td>60.87</td>
<td>4.64</td>
<td>.01</td>
</tr>
<tr>
<td>Reading</td>
<td>87.50</td>
<td>63.64</td>
<td>6.14</td>
<td>.01</td>
</tr>
<tr>
<td>Writing–copy</td>
<td>80.00</td>
<td>65.00</td>
<td>5.40</td>
<td>.02</td>
</tr>
<tr>
<td>Writing–dictation</td>
<td>76.92</td>
<td>70.59</td>
<td>6.65</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note. VMI = Developmental Test of Visual-Motor Integration.
many of the children in the at-risk group received occupational therapy or physical therapy in the intervening years, their motor performance might have improved as a result of treatment efforts over time.

The results of our study indicated that no significant differences existed in the Conners ASQ between the groups at follow-up. Because the Conners ASQ taps into parental perceptions of the attention and hyperactive behaviors of their children, the results suggest that the MAP is not sensitive to these areas. These findings may be explained by noting that the MAP is a "child-friendly" assessment. The different items are presented in a playful fashion and are devised to be executed in short periods of time, up to a few minutes each, so that even children with short attention spans can demonstrate their optimal abilities. Given the importance of the behavioral issues in academic achievement, the results of this study might indicate that the MAP should be reinforced with a complementary test (e.g., an index of hyperactivity) that refers specifically to this variable.

Another limitation of this study is sample size. Power estimates indicated that had a larger sample size been used, the VMI, LOTCA, HHE (copy and dictation), and reading proficiency scores would have discriminated between the groups even more. In contrast, the results of those tests for which group scores did not reach significant differences (BOTMP, Conners ASQ, VADS) would not have been affected by increasing the number of participants. These findings support the reliability of our current study results despite the small sample size.

To achieve optimal predictability, tests that intend to screen for developmental risk during a presymptomatic period must be sensitive to participants who demonstrate later developmental difficulties (Meisels & Wasik, 1990). Sensitivity and specificity indexes were calculated for the MAP with respect to each one of the outcome criteria variables that demonstrated statistical differences between the groups. The highest coherence was between the MAP and the VMI and the reading and writing proficiency tests. These findings are congruent with the MAP objectives of predicting academic performance (Miller, 1988a, 1988b, 1988c; Miller et al., 1987).

The use of a long prediction interval (5–7 years) is an important new issue addressed by this study. Although literature on young children’s screening tools frequently calls for long-term longitudinal research, few scales actually demonstrate this attribute (Badian, 1988; Gredler, 1997). The heterogeneous nature of learning disabilities makes their long-time prediction more challenging; shorter periods for prediction increase the likelihood that predictive statistics will appear high. A critical problem with short-term follow-up is that the cognitive levels of proficiency demonstrated by young children are discontinuous when compared with later academic tasks. Different cognitive tasks represent relevant markers for academic success at later ages. Thus, the weaknesses that appear in earlier ages may be expressed in cognitive processes at a later age. In addition, no academic difficulties may be evidenced until children are required to address more complex or abstract subjects in their curriculum (Lavin, 1995). A relatively long time lapse, such as that seen in this study, increases the chance of uncontrolled-for effects of history; intervening experiences may weaken the association between screening outcomes and follow-up criteria (Miller, 1988a, 1988b). Perhaps, the inclusion of verbal and nonverbal cognitive items in the MAP that demand memory and problem-solving skills have given it the predictive power to relate to aspects of future cognitive; perceptual-motor constructs; and academic performance, such as reading and handwriting proficiency.

**Conclusion**

The results of this study strengthen the assertion that the use of the MAP can be very beneficial in clarifying potential sources of observed deficits in important daily tasks (Coster, 1998; Dunn, 2000b). Significant differences were demonstrated in school performance between the two groups of children (at risk, not at risk) as designated by the original MAP results across a substantial period of time (5–7 years). These results provide evidence of the strong predictive capability of the MAP in accurate prediction of academic performance. In addition, the examination of academic and varied performance components in another cultural context allows for better insight into the aspects of performance that impinge on the occupational performance of preschool-age children irrespective of cultural context.

**Acknowledgments**

We thank Beit Issie Shapiro for its cooperation in providing access to the original MAP screening data that formed the basis for our follow-up findings. We especially acknowledge the assistance of Dana Roth, PhD, Director of Research and Development at Beit Issie Shapiro, Raanana, Israel, and Michele Shapiro, MSc, OT, who initiated and conducted the difficult, yet important task of screening the preschool children of Raanana, Israel.

This study was done in partial fulfillment of the second author’s requirement for a master’s degree in occupational therapy from the School of Occupational Therapy, The Hebrew University–Hadassa Medical School.
References


