Use of the Bruininks-Oseretsky Test of Motor Proficiency in Occupational Therapy

Brenda N. Wilson, Helene J. Polatajko, Bonnie J. Kaplan, Peter Faris

Key Words: assessment process, occupational therapy • motor skills

Objective. In addition to the need for good measurement tools in occupational therapy, there is a need for the tools to be used knowledgeably. The purpose of this article is to investigate the usefulness of the Bruininks-Oseretsky Test of Motor Proficiency (BOTMP) for both descriptive (diagnostic) and evaluative (change over time) purposes.

Method. The typical profile of subtest scores for children with mild motor problems revealed that certain subtests of the BOTMP may be better indicators of motor problems for these children than others. An analysis also was performed to compare the use of raw (point) scores with standard (age-adjusted) scores in evaluating change.

Results. Four subtests that provide a greater degree of discrimination between children with and without motor problems were identified. Raw (points) scores were found to provide a more valid measure of change over time than standard (age-related) scores.

Conclusion. We recommend that, for clinical use, the BOTMP subtest standard scores be used for diagnostic purposes and that the raw scores be used for evaluative purposes.

T

here is no question of the need for good measurement tools in occupational therapy (Miller, 1989) and of the even greater need to use the tools knowledgeably. King-Thomas and Hacker (1987) stated that

the importance of professional accountability in pediatric assessment should never be underestimated. Each person participating in selection and application of standardized or nonstandardized screening or diagnostic tests must be constantly aware of the potential effects of the decisions they will be making. Children’s lives can be positively or negatively affected by the results, depending on the validity and interpretation of findings. (pp. 9–10)

Assessment by occupational therapists is a process whereby therapists evaluate clients’ actual performance of daily life tasks. Fisher (1992a) described this process as follows: “(a) the daily life problem is defined; (b) the interrelationship among occupational performance and underlying musculoskeletal, neurologic, cardiopulmonary, or cognitive capacities is evaluated . . . Our concern with the prerequisite neuromotor, psychosocial, and cognitive-perceptual performance capacities is framed in their impact on occupational performance” (p. 184). The functional problems that children demonstrate at home, in school, and at play are generally defined through observations, interviews, and criterion-referenced measures. Subsequently, a therapist may look for the underlying, prerequisite capacities that can negatively influence the child’s occupational performance. This stage of the assessment process is done with both standardized and nonstandardized tools, both of which play an important part in assessment.

Developing the framework of functional limitations and occupational competency can be accomplished with
observational and clinical judgment. However, in evaluation of specific components of occupational performance, formal standardized assessments allow uniformity and objective administration and scoring, and thereby increase the likelihood that a test gives a true measure of the component being evaluated. Standardized assessments have the potential to objectify our clinical judgments. One such test used widely in occupational therapy is the Bruininks-Oseretsky Test of Motor Proficiency (BOTMP) (Bruininks, 1978), a pediatric test of fine and gross motor skills. Among one sample of pediatric occupational therapists (n = 40), 92.6% used the BOTMP in their practice (Yack, 1989). In another survey taken of 20 Ontario, Canada, treatment centers, the BOTMP was one of the five most commonly used measures of motor function (Gowland et al., 1991). The BOTMP has enjoyed such wide clinical acceptance because it measures skills important to pediatric practice, because it is perceived to have good psychometric properties, and because few other tests of motor skills exist for the school-aged child.

The BOTMP is used to assess children, aged 4½ to 14½ years, who demonstrate motor problems not related to obvious neurological disorders. These children may be diagnosed as having developmental coordination disorder (DCD) (American Psychiatric Association, 1987), sensory integrative problems (Ayres, 1972), or clumsy child syndrome (Losse et al., 1991). In this paper, we will use the term mild motor problems to include all of the above diagnostic frameworks.

In developing a guide to help clinicians evaluate the usefulness of measurement tools, Law (1987) categorized instruments on the basis of their purpose, as originally outlined by Kirshner and Guyatt (1985). Two of the categories are descriptive measures, which are used diagnostically to describe persons and to characterize the differences between persons on the construct being measured, and evaluative measures, which are used to quantitatively changes in a person over time and which must be sensitive to those changes.

The purpose of this article is to explore the usefulness of the BOTMP for children with mild motor problems in order to assist clinicians in making more meaningful use of the test as a descriptive and evaluative instrument. To illustrate the use of the BOTMP as a descriptive instrument, a typical profile of subtest scores for children with mild motor problems from three studies will be presented. Then, to illustrate the use of the BOTMP as an evaluative instrument, a secondary analysis of data from a recently published study will be used to compare the composite scores to the point scores of individual subtests.

**Review of Psychometric Properties**

**Description of the BOTMP**

The BOTMP was developed "to provide educators, clinicians and researchers with useful information to assist them in assessing the motor skills of individual students, in developing and evaluating motor training programs, and in assessing serious motor dysfunctions and developmental handicaps" (Bruininks, 1978, p. 11). It is a 46-item test, composed of eight subtests (see Table 1). Four of these subtests are combined to provide the Gross Motor Composite Score, and three subtests are combined to provide the Fine Motor Composite Score. All of the subtests, in combination, provide a total picture of a child’s motor performance.

In a review of the BOTMP, King-Thomas and Hacker (1987) examined test-retest reliability on a sample of 63 second graders and 63 sixth graders, all without disabilities. The test-retest reliability of the coefficients ranged from .68 to .88. (Qualifications of the raters were not given in the test manual.) Coefficients above .80 were reported for all subtests except Balance (subtest 2) and Response Speed (subtest 6), which suggests that these subtests should be interpreted cautiously (King-Thomas & Hacker, 1987). They also found that the fine motor composite is more reliable for children in Grade 2 than in Grade 6, whereas the gross motor composite is more reliable for children in Grade 6 than in Grade 2. Because better reliability was obtained for the total test and short form than for individual subtests, King-Thomas and Hacker (1987) did not recommend basing decisions about the child on the subtest scores alone.

Hattie and Edwards (1987) had a less positive review of the BOTMP. They have questioned both the dependability of the test scores and its use with children who have motor or intellectual problems. Their review described studies that found that gender differences existed, item consistency was low, standard errors of measurement were high, and grouping of subtests had limited usefulness. They provided summary comments about each of the subtests to aid in interpretation and concluded that the "test may be of use as a diagnostic tool if the pattern of items or subtests are carefully inspected to ascertain what items a child finds easy or difficult" (p. 111).

Interrater reliability has been evaluated only for

---

**Table 1**

<table>
<thead>
<tr>
<th>Composite</th>
<th>Subtest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Motor Composite</td>
<td>1 = Running Speed and Agility (RSA)</td>
</tr>
<tr>
<td></td>
<td>2 = Balance (Ba)</td>
</tr>
<tr>
<td></td>
<td>3 = Bilateral Coordination (BC)</td>
</tr>
<tr>
<td></td>
<td>4 = Strength (St)</td>
</tr>
<tr>
<td>Battery Composite</td>
<td>5 = Upper-Limb Coordination (ULC)</td>
</tr>
<tr>
<td>Fine Motor Composite</td>
<td>6 = Response Speed (RS)</td>
</tr>
<tr>
<td></td>
<td>7 = Visual-Motor Control (VMC)</td>
</tr>
<tr>
<td></td>
<td>8 = Upper-Limb Speed &amp; Dexterity (ULSD)</td>
</tr>
</tbody>
</table>

*Note: BOTMP = Bruininks-Oseret,sky Test of Motor Proficiency*
Visual-Motor Control (subtest 7) and was reported to range from .77 to .97 for individual test items (Bruininks, 1978). Correlations for the total subtest score reached .96 for raters who had received training and .90 for those who used only the scoring instructions in the manual. Lack of any examination of interrater reliability on the other seven subtests indicates a major limitation of the BOTMP and suggests the need to use the same rater when the test is being readministered. In addition, because the sample data for reliability and validity measures were gathered from children without motor problems, Gowland et al. (1991) have recommended caution in assuming that these measures would be the same for children with motor problems.

The manual for the BOTMP (Bruininks, 1978) reported the standard errors of measurement (SEM) for subtests and for composite scores. The SEM for the subtests, which have a mean of 15 and a standard deviation of 5, is 2 or 3 standard score points. For the composite scores, which have a mean of 50 and a standard deviation of 10, the SEM is 4 or 5 standard points. Cunningham Amundson and Crowe (1993) have recommended applying two SEMs to achieve a 95th confidence band. For the BOTMP, this is equivalent to one standard deviation. Hattie and Edwards (1987) found this degree of difference in scores to be very high and noted that it was especially problematic for Running Speed and Agility (subtest 1), Balance (subtest 2), Upper Limb Coordination (subtest 5), and Response Speed (subtest 6).

To measure internal consistency, the individual items were correlated with their subtest scores and with total test scores. Moderate to high correlations were reported (Bruininks, 1978), indicating that the functions being measured within each subtest are related, or relatively homogeneous. These correlations were stronger for the younger age groups than for older children and teenagers. A factor analysis was completed at the time the test was constructed to lend support to the groupings of the test items into the various subtests. Although results of this analysis supported the groupings of upper limb coordination, balance, strength, and bilateral coordination, most of the items in the fine motor subtests clustered together with other gross motor items in a grouping that appeared to reflect general motor development.

Three studies that support construct validity are cited in the manual (Bruininks, 1978). They contrast the performance of children with and without mental or learning disabilities. Only Response Speed (subtest 6) did not discriminate between these two groups of children. In addition, content validity was established by relating relevant research studies of motor proficiency with the content of the BOTMP.

Bruininks (1978) outlined the process of developing norms for the BOTMP, which began with the transformation of the item raw scores (e.g., time taken to complete the task, number of errors) into a set of scaled scores called point scores. These point scores were then normalized to produce the subtest standard scores, which have a mean of 15 and a standard deviation of 5, for each age group. The standard score norm lines for each age group were plotted and smoothed, and norm lines for each 6-month age group were calculated by interpolation. Gross motor and fine motor composite standard scores were constructed with the summation of the subtest standard scores, and these scores were transformed to normalized standard scores that have the same relative meaning for each age group. These normalized composite standard scores, which were further smoothed in their development, have a mean of 50 and standard deviation of 10. From these scores, percentiles ranks, stanines, and age equivalent scores were calculated. Although transforming and smoothing the test scores were necessary to provide normative information and establish a clear relationship between floor and ceiling, it should be acknowledged that, in some instances, transformations reduce the clarity of the scores.

It would appear from this review of reliability and validity that the BOTMP is an appropriate descriptive measurement tool of the motor abilities of children, within the limits of undefined interrater reliability. However, questions remain about its use as an evaluative tool.

The BOTMP is often used as an evaluative instrument to measure changes in gross and fine motor skills in children receiving occupational therapy (Clark, Mailoux, Parham, & Bissell, 1989; Gowland et al., 1991; Humphries, Wright, McDougall, & Vertes, 1990; Humphries, Wright, Snider, & McDougall, 1992; Law, 1987; Polatajko, Law, Miller, Schaffer, & MacNab, 1991; Wilson & Kaplan, in press; Wilson, Kaplan, Fellowes, Gruchy, & Paris, 1992). Ziviani, Poilusen, and O'Brien (1982) correlated the BOTMP scores with the Southern California Sensory Integration Tests (Ayres, 1972) scores of children with learning disabilities. From the moderate but significant results, they concluded that the BOTMP could be used as a screening test for children with possible sensory integrative dysfunction. They also concluded that the test “may provide an objective assessment of motor changes” (p. 525). This conclusion is questionable considering that the “most appropriate characteristics included in an evaluative instrument are those that can be shown to be sensitive to change within an individual. Extraneous information in this type of instrument leads to a decreased responsiveness” (Law, 1987, p. 134). The usual indicators of validity for descriptive and predictive measures are not necessarily useful in determining the validity of an evaluative measure. The only valid test of an evaluative measure is a consistent correlation between the instrument in question and another measure of change that is known to have taken place (Law et al., 1989). A review of the literature did not reveal any studies that specifically examined the sensitivity of the BOTMP or its ability to measure response to treatment.
Clinical Use of the BOTMP

The following analyses were designed to help determine the appropriateness of using the BOTMP as a descriptive and as an evaluative instrument.

Data Source

The data used in this review came from three samples: 105 children with developmental coordination disorder (DCD) (American Psychiatric Association, 1987) and sensory integrative dysfunction (SID) (Ayres, 1972) who participated in two treatment efficacy studies (Polatajko et al., 1991, Wilson et al., 1992), and 15 children with learning disabilities and coordination-clumsiness problems (Horak, Shumway-Cook, Crowe, & Black, 1988).

The first two samples, aged 5 years 2 months to 8 years 6 months ($M = 6$ years 8 months; $SD = 12.5$ months), would best be categorized as having DCD (American Psychiatric Association, 1987); that is, the children had coordination and clumsiness problems that interfered with their performance in academics and activities of daily living. All had been referred to physical therapy or occupational therapy for these problems and, upon further occupational therapy assessment, were diagnosed with sensory integrative dysfunction. They were all of average intelligence, defined as a score of 85 or more on the Wechsler Intelligence Scale for Children (WISC) in the Wilson et al. (1992) study and as a score of 80 or more on the Slosson Intelligence Test in the Polatajko et al. (1991) study. The academic performance of this group of children was delayed by 6 to 12 months on the Woodcock-Johnson Psychological Battery (WJPB) (Polatajko et al., 1991) or was below average on at least one of three subtests of the WJPB (Wilson et al., 1992). Schaffer, Law, Polatajko, and Miller (1989) examined the performance of the Polatajko et al. (1991) sample ($n = 76$) and found that 80% of the subjects were delayed in reading, and 91% were delayed in written language. Exclusion criteria for the samples in both of these studies included medical problems that interfered with treatment or neurological disorders, such as epilepsy or cerebral palsy. In the Wilson et al. (1992) study ($n = 29$), children with Attention Deficit Disorder were also excluded. Results showed that subjects in Grades 2 and higher experienced learning problems, whereas younger subjects exhibited many characteristics of preacademic problems and were considered to be at risk for learning disabilities. More detailed descriptions of these subjects are available in Polatajko et al. (1991) and Wilson et al. (1992).

The third study sample (Horak et al., 1988) was recruited from an outpatient program for children with developmental disabilities and had a mean age of 9 years 4 months ($\pm 1.5$ years). Children in this sample were at least 2 years delayed in two or more academic subjects, which was not consistent with their estimated intelligence. There were no overt auditory, visual, or neurological deficits. No information on IQ was reported.

The BOTMP as a Descriptive Instrument

Although a number of studies report the use of the BOTMP with learning disabled populations, most studies simply report the composite scores. In looking for a profile of motor scores from the BOTMP subtests for children with mild motor problems, we found one source in the literature where subtest scores were actually reported, although only for the gross motor subtests (Horak et al., 1988). The BOTMP subtest standard scores from this study were plotted, together with data available from the Polatajko et al. (1991) study and the Wilson et al. (1992) study to determine whether a profile could be detected. The samples seemed to have a somewhat similar pattern of subtest standard scores (see Figure 1). Furthermore, they seemed to differ from test means for a number of subtests. There were only two subtests in which all of the mean standard scores fell near the test mean or within one standard deviation of the test mean, and in no subtest did any sample score above the test mean.

The profile was established by visual inspection of Figure 1. To categorize the profile, the following three decision rules were used:

1. Subtests in which all of the sample means were within one standard deviation of the mean, were considered to represent an area of normal function for children with learning disabilities and motor problems. These subtests included Strength (4) and Response Speed (6).
2. Subtests in which all of the sample means were more than one standard deviation below the mean were considered problematic. These subtests included Running Speed and Agility (1), Balance (2), Visual Motor Control (7), and Upper Limb Speed and Dexterity (8).
3. If the pattern was mixed, with some below and some within one standard deviation, the subtest was considered to represent an area of questionable performance. These subtests included Bilateral Coordination (3) and Upper Limb Coordination (5).

Figure 1 shows that children with learning disabilities and motor problems performed within normal limits on Bilateral Coordination (subtest 3), Strength (subtest 4), Upper Limb Coordination (subtest 5), and Response Speed (subtest 6). They had difficulty with Running Speed and Agility (subtest 1), Balance (subtest 2), Visual-Motor Control (subtest 7), and Upper Limb Speed and Dexterity (subtest 8).

This profile has implications for the identification of children for treatment programs and for the identification of treatment strategies.
of treatment goals. The four subtests that have been identified as problematic may be the best ones to discriminate between children with and without mild motor problems. Thus, in addition to overall motor performance as reflected by the test composite scores, performance on these subtests should be specifically examined for possible poor performance. Findings may add further support to the diagnosis of mild motor problems when the test is used descriptively.

Analysis of the BOTMP as an Evaluative Instrument

As previously discussed, there is little support for the use of the BOTMP to measure change over time, although the same is true for most tests used in therapy and education. However, the BOTMP is commonly used in clinical practice and in research as an evaluative measure. The author of the test has recommended that it is valid to readminister the assessment within a 3-4 week period (R. H. Bruininks, personal communication, February 9, 1984) and that the test was developed to evaluate the effectiveness of motor programs (Bruininks, 1978).

With respect to a child who has problems with motor performance and who is receiving treatment, readministration of the BOTMP may confirm a clinical impression of improvement. However, if the progress is slow, the normative scores (the subtest standard scores and composite scores) may not show the change, perhaps because the normative scores on the BOTMP are based on a sample of children without motor delays. These scores may therefore show progress only if the rate of change is faster than normal maturation, or if the child did not change age groups between pretest and posttest. Indeed, if progress is occurring at a rate slower than normal maturation, the normative standard scores may actually decrease. A lower score could prove very demoralizing for the child, the
family, and the therapist. To use a score that may indicate progress only if it changed faster than maturation seems overly conservative (Morris, 1989).

It may be more appropriate to report any progress that a child makes in subtest point scores. As Guyatt, Walter, and Norman (1987) pointed out, "one problem . . . is that for newly developed instruments (and for a distressingly large proportion of established ones as well) we do not know the change in score that constitutes a clinically important difference" (pp. 175–176). The purpose of retesting should be to determine whether the child’s skill level is improving, by comparing the child’s performance before and after treatment. For treatment outcome purposes, it is far more useful to compare the child’s performance to his or her previous performance than to that of the normative sample.

Thus, care must be taken in interpreting a child’s posttest normative scores (standard or composite). No change should not be interpreted as no improvement in the child. Where there is no change in the normative scores, the subtest point scores should be examined for change. Although one might anticipate that the subtest point scores and standard scores would indicate the same outcome, this is not necessarily the case. To illustrate the differences in change that the BOTMP subtest standard scores and point scores may indicate, the individual scores of four subjects (subjects 19, 5, 29, and 18) in the Wilson et al. (1992) sample were plotted on graphs (see Figures 2 through 5). The direction of change between pretest and posttest for each subject was presented as standard scores and point scores. In the case of subject 5 (see Figure 3), the direction of change in performance indicated by the standard score for most subtests is mirrored by the point score. However, at times, the subtest standard score and point score indicate different outcomes. The point score indicates that the child may have improved, although the standard score shows little or no change (see Figure 4, subtest 4 and Figure 2, subtest 8). In some cases, the standard score even shows deterioration (see Figure 5, subtest 7, and Figures 4 and 5, subtest 2). It is even more important not to interpret a deterioration on the normative score at face value. Again the subtest

![Figure 2](http://ajot.aota.org/10/24/2018) Terms of Use: http://AOTA.org/terms
point scores should be examined. The subtest standard scores could indicate deterioration when, as in the aforementioned examples, the point scores show that the child actually performed better. Figures 2 through 5 show a number of scenarios depicting the agreement and lack of agreement between subtest point scores and the age-corrected subtest standard scores. We investigated 14 subjects' performance on eight subtests and found that 25% of the standard scores did not agree with the direction of change of the raw scores. We found the least amount of disagreement on Visual-Motor Control (subtest 7) and the most on Bilateral Coordination (subtest 3).

Another way of analyzing the usefulness of the BOTMP as an evaluative tool is to examine how statistical analyses change depending on the type of score used. Table 2 compares the results of MANOVAs run on the raw versus the standard subtest scores of the BOTMP for the Wilson et al. (1992) study. Because the standard score adjusts for age, and age was significantly related to subtest scores, age was used as a covariate in the analysis of the raw score. Even though age (developmental change) was accounted for in both analyses, the F ratios tended to be larger when the analysis used raw scores, especially in the

### Table 2
Comparison of Analysis Using Raw Scores and Standard Scores

<table>
<thead>
<tr>
<th>Subtest of BOTMP</th>
<th>Pretest to Midtest</th>
<th>Pretest to Midtest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw Score</td>
<td>Standard Score</td>
</tr>
<tr>
<td></td>
<td>F (df)</td>
<td>F (df)</td>
</tr>
<tr>
<td>Gross Motor Subtests</td>
<td>.55 (4,24)</td>
<td>1.62 (4,24)</td>
</tr>
<tr>
<td></td>
<td>p = .70</td>
<td>p = .20</td>
</tr>
<tr>
<td>Upper Limb Coordination Subtest</td>
<td>7.25 (1,26)</td>
<td>3.88 (1,27)</td>
</tr>
<tr>
<td></td>
<td>p = .01**</td>
<td>p = .08</td>
</tr>
<tr>
<td>Fine Motor Subtests</td>
<td>5.08 (3,25)</td>
<td>1.97 (3,29)</td>
</tr>
<tr>
<td></td>
<td>p = .05*</td>
<td>p = .15</td>
</tr>
</tbody>
</table>

*p < .05

**p < .01
Fine Motor subtests and Upper Limb Coordination subtest. This finding is probably related to the smoothing and transforming of the standard scores.

In conclusion, when measuring the progress of children, it is more useful to compare a child's performance to his or her previous performance (with point scores) than to compare the performance to that of the normative sample (with the standard scores). The standard scores will only reflect change that is faster than typical maturation, which is a rate of progress that few children with mild motor problems are able to achieve. The application of SEMs may assist in differentiating real progress from measurement error. Although standardized tests are widely used to measure progress and study treatment efficacy, other measurement methods, such as functional performance outcomes, may provide useful information.

Discussion
On the basis of the analyses presented here, we recommend that occupational therapists be aware of the limitations in the use of normalized, standard scores when using the BOTMP for descriptive or diagnostic purposes. Composite scores, whether gross motor, fine motor, or battery composite, are often necessary to use in an assessment requiring standardized measures. However, as accountable therapists, examiners should also look carefully at subtest scores and be cognizant of which subtests are shown to be the best discriminators of relative strengths and deficits in motor skills.

From this analysis and review of the development of the BOTMP, it appears that the Running Speed and Agility, Balance, Visual Motor Control, and Upper Limb Speed and Dexterity subtests are likely to provide the greatest degree of discrimination between children with and without motor problems. The remaining four subtests are least likely to discriminate between children with and without motor problems, with the performance on Response Speed being the least useful.

When the BOTMP is used for evaluative purposes, it is recommended that the subtest point scores be used, rather than the normative scores (standard and composite scores) when using the BOTMP for descriptive or diagnostic purposes. Composite scores, whether gross motor, fine motor, or battery composite, are often necessary to use in an assessment requiring standardized measures. However, as accountable therapists, examiners should also look carefully at subtest scores and be cognizant of which subtests are shown to be the best discriminators of relative strengths and deficits in motor skills.

Figure 4. Agreement between standard score and point score change profile for Subject 29. RSA = Running Speed and Agility, Ba = Balance, BC = Bilateral Coordination, St = Strength, ULC = Upper Limb Coordination, RS = Response Speed, VMC = Visual-Motor Coordination, ULSD = Upper Limb Speed and Dexterity.
Figure 5. Agreement between standard score and point score change profile for Subject 18. RSA = Running Speed and Agility, Ba = Balance, BC = Bilateral Coordination, St = Strength, ULC = Upper Limb Coordination, RS = Response Speed, VMC = Visual-Motor Coordination, ULSD = Upper Limb Speed and Dexterity.

The use of subtest point scores will result in a more precise measurement of function, because gains or deterioration will be related to specific areas of motor control. In addition, scores that have undergone statistical transformations will be less exact in their ability to detect real changes that occurred. Because these standard scores are age adjusted, progress will not be reflected in the test scores unless the progress is faster than typical maturation (which is not likely to occur with children who have motor problems). Therapists should consider using the subtest point scores as a more accurate measure of change. One way to measure clinical change is to establish an initial baseline performance (e.g., test the child before placing on waiting list, or use the interval between screening and the start of treatment as baseline), and then compare the improvement after treatment to the growth that occurred during a no-treatment baseline period.

It is also important to be aware that the interrater reliability for seven of the eight BOTMP subtests has not been established. Although intrarater reliability has never been evaluated, it is assumed to be higher than interrater reliability. Therefore, we recommend that the same examiner provide the initial assessment and the reevaluation of the child. If the same examiner is not used for reassessment, comparison of before and after scores cannot be made with the same confidence.

This article has outlined the limitations of a commonly used pediatric test. The limitations of the BOTMP are not unlike those of many standardized assessments used in occupational therapy. It is important that therapists be aware of the limitations and be clear "about what we will know given the results of a test, why we want to know it, and whether it is what we really want to know" (Fisher, 1992b, p. 280).

Implications for Future Research

There is a need to establish interrater and intrarater reli-
ability for each of the BOTMP subtests, as well as for the composite scores. A measure cannot be used with complete confidence unless its reliability is well established.

It is also apparent that the sensitivity of the BOTMP has never been established through the conventional method of comparing it with another measure that is known to accurately detect change over time. One of the obvious reasons for this lack of comparison is the dearth of evaluative instruments in this area. Of the 13 other sensorimotor tests reviewed by De Gangi (1987), none would be appropriate for comparison with the BOTMP for evaluative purposes, as none themselves have been proven as evaluative measures. In the absence of a gold standard, a criterion measure available for comparison with the BOTMP, it will be necessary to use this test very cautiously when evaluating children over time.

The usefulness of the BOTMP to diagnose and describe the motor problems of children is an important issue for occupational therapists to address. Reviews of the test are mixed, and the limitations reported here must be considered further if the test is to be used with confidence.

Acknowledgments
We thank Doug Savage for graphics design, and Anne Campbell, OT(C), and anonymous reviewers for their review of earlier drafts of this manuscript. During the analysis of this data, Dr. Polatajko was supported by the Alberta Mental Health Research Fund as a Visiting Scientist. We also acknowledge the financial support of the Alberta Children’s Hospital Foundation.

References