Objective. This study examined the equivalence reliability of two administration methods for the Functional Independence Measure for Children (WeeFIM®). The two methods were direct observation of child performance and parental interview.

Method. Thirty children between the ages of 19 months and 71 months with identified developmental disabilities were included in this study. The direct observation and interview methods were administered in random order to each subject within a 3-week period. The direct observation was completed in the educational setting, and the WeeFIM interview was obtained by either in-person interview or telephone interview with the parent.

Results. The intraclass correlation coefficient (ICC) for total WeeFIM ratings was .93, indicating consistency between the two administration methods. Analysis of the WeeFIM Motor (13 items) and Cognitive (5 items) domains yielded ICC values of .93 and .75, respectively, suggesting greater consistency for items measuring motor-related skills.

Conclusion. The results demonstrate good agreement for total ratings when the WeeFIM is administered by direct observation and by interview with a parent. The findings establish the clinical usefulness of information collected by clinical interview.

Historically, pediatric assessments have consisted of developmental scales designed to determine behavioral status in motor, cognitive, and social skill areas (Bundy, 1990; Dunn, 1993; Fisher, 1992a, 1992b; Haley, Hallenborg, & Gans, 1989). Concerns about the appropriateness of these scales, especially for children with disabilities, have recently been raised (Coster & Haley, 1992). Developmental scales often suggest an invariant order of skill attainment (Garwood, 1982; Guess & Noonan, 1982). In reality, there are variations of the developmental progression, and it is common for children to demonstrate developmentally higher level skills before achieving lower level abilities (e.g., a child who walks without learning to crawl). Therefore, typical developmental assessments often identify and describe milestones that may not be appropriate for children with disabilities (Garwood, 1982).

Recently, evaluation of functional skills and functionally focused intervention have become priorities in early intervention and developmental research (Coster & Haley, 1992; Marosszeky, 1993). Functionally focused treatment and educational outcomes are often necessary.

Patricia A. Sperle, Kenneth J. Ottenbacher, Susan L. Braun, Shelly J. Lane, Susan Nochajski

Key Words: activities of daily living evaluation • evaluation • measurement scales
for program planning, program evaluation, documentation, quality care, and reimbursement (Haley, 1993; Msall et al., 1994).

**Functional Evaluation**

Functional evaluation in children has been described as “an effort to systematically describe and measure a child’s abilities and limitations when performing the activities of daily living” (McCabe & Granger, 1990, p. 121). Functional evaluation is an important component of a comprehensive evaluation because it identifies what the child can actually do within a specific environment. Several authors have described the conceptual advantages of functional evaluation over traditional developmental approaches (Haley et al., 1989; Msall et al., 1994). First, the observation of functional performance allows the use of special equipment or assistive devices. Second, emphasis on functional skills is more relevant to a child with physical impairment than the achievement of many developmental milestones, such as skipping or standing on one leg. Third, the endpoint of functional outcome is viewed as more important than the method used to achieve the result. For instance, functional locomotion can be achieved via either walking or using a wheelchair.

Two pediatric functional assessments currently exist for use with children: the Functional Independence Measure for Children (WeeFIM®) (Guide for the Functional Independence Measure for Children, 1993) and the Pediatric Evaluation of Disability Inventory (Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 1992). Each instrument measures functional ability, taking into account the use of special equipment and the amount of caregiver assistance. The ultimate goal of the WeeFIM is to “measure changes in function over time to weigh the burden of care in terms of physical, technologic, and financial resources” (Braun & Granger, 1991, p. 41).

The WeeFIM is designed to be administered by either direct observation or interview of a primary caregiver (parent or teacher) who knows the child well. The equivalence reliability of the two administration methods has not been previously studied. Portney and Watkins (1993) stated that establishing equivalence reliability is “necessary if absolute values are to be compared or equated across tests and to generalize findings from one study to another or from research to practice” (p. 62). The purpose of the present investigation was to compare the ratings of the two methods, using the intraclass correlation coefficient (ICC), in a sample of children with disabilities.

**Method**

**Subjects**

The subjects were 30 children (17 girls, 13 boys) who received special services for identified developmental disabilities. The extent of the subjects’ disabilities ranged from mild to severe, multiple disability. Their ages ranged from 19 months to 71 months ($M = 48.0$ months, $SD = 13.0$ months). Six subjects were included in the 12-month to 36-month group, 17 in the 37-month to 60-month group, and 7 in the 61-month to 84-month group. Most subjects came from lower- to upper-middle class backgrounds, but socioeconomic status was not specifically examined. The predominant medical condition of the subjects was cerebral palsy ($n = 12$). Other conditions included Down’s syndrome ($n = 6$), developmental delay ($n = 7$), and congenital abnormalities ($n = 5$).

The convenience sample was recruited from three educational facilities in western New York state. All three facilities have educational day programs designed specifically for children with disabilities. Many of the subjects were mainstreamed into day programs that included children without disabilities. All three facilities involved integration of children without disabilities to some extent. Most of the subjects received related services (i.e., occupational therapy, speech therapy, physical therapy) as part of their individualized education programs or individualized family service plans.

**Instrument**

The WeeFIM is a pediatric functional assessment developed in 1987 by the Uniform Data System for Medical Rehabilitation (UDS) at the State University of New York (S.U.N.Y.) at Buffalo. It is an adaptation of the Functional Independence Measure (FIM), which is designed to measure severity of disability in adults (Guide for the Uniform Data System for Medical Rehabilitation, 1993). The WeeFIM was developed to assess and track functional abilities in children ages 6 months to 7 years across health, developmental, educational, and community settings (Msall et al., 1994). Key characteristics of the WeeFIM are the use of a minimal data set, emphasis on actual performance, and the ability to be used by different disciplines (Msall, DiGaudio, & Duffy, 1993).

The WeeFIM (see Figure 1) contains 18 items divided over six subscales: Self-Care, sphincter control, transfers, locomotion, communication, and social cognition. Each subscale consists of 2 to 6 items that are scored separately. An ordinal rating system ranging from 7 (complete independence) to 1 (total assistance) is used. A rating from 5 to 1 indicates that the child requires some level of assistance from another person to complete the activity. A rating of 7 or 6 means that the child can complete the activity independently but may require an assistive device, need more than a reasonable amount of time, or need assistance when safety is a concern. No zero or non-applicable ratings can be given. The minimum possible total rating is 18 (total dependence in all skills), and the
The WeeFIM can be administered either through direct observation, interview, or a combination of observation and interview. A variety of professionals can use the WeeFIM, but training is recommended to ensure appropriate administration and scoring (Msall, Braun, & Granger, 1990). Validity and interrater reliability have been consistently observed in the school setting because these items are not skills typically performed in school (see Figure 1). These items were observed and rated when possible, but because they were not consistently rated for all subjects, they were not included in the analysis.

Twenty-seven parents were contacted by phone for the interview administration of the WeeFIM, and three interviews were completed in person. A previous study has demonstrated reliability between the ratings obtained by in-person and telephone interview methods (Taylor et al., in press). All items were rated in the interview administration.

The observation and interview administrations were done in random order and were completed within a period of 14 days for 27 subjects. Absenteeism from school and inability to reach parents by phone increased the time span to 21 days for the 3 remaining subjects.

Data Analysis

Paired t tests were used to check for significant differences between mean scores obtained by direct observation versus mean scores derived from parent interview. Separate paired t tests were conducted for items, subscale scores, and the mean Motor and Cognitive domain scores. The ICC was computed to determine agreement among ratings. The ICC is a reliability coefficient calculated from estimates obtained through analysis of variance and specifically examines agreement among raters (Portney & Watkins, 1993). The ICC model (2,1) (Shrout & Fleiss, 1979) was computed in the current investigation because it includes variance components related to method of administration and provides a conservative estimate of agreement (Portney & Watkins, 1993).

The relation of scores between the two administra-

---

**Figure 1.** Sample WeeFIM rating form. Note: From *Guide for the Functional Independence Measure for Children (WeeFIM) of the Uniform Data System for Medical Rehabilitation. Version 4.0—Community/Outpatient*, 1993, Buffalo, NY: State University of New York at Buffalo. Copyright 1993 by State University of New York at Buffalo. Adapted with permission.

maximum is 126 (complete independence in all skills).

The WeeFIM can be administered either through direct observation, interview, or a combination of observation and interview. A variety of professionals can use the WeeFIM, but training is recommended to ensure appropriate administration and scoring (Msall, Braun, & Granger, 1990). Validity and interrater reliability have been consistently observed in the school setting because these items are not skills typically performed in school (see Figure 1). These items were observed and rated when possible, but because they were not consistently rated for all subjects, they were not included in the analysis.

Twenty-seven parents were contacted by phone for the interview administration of the WeeFIM, and three interviews were completed in person. A previous study has demonstrated reliability between the ratings obtained by in-person and telephone interview methods (Taylor et al., in press). All items were rated in the interview administration.

The observation and interview administrations were done in random order and were completed within a period of 14 days for 27 subjects. Absenteeism from school and inability to reach parents by phone increased the time span to 21 days for the 3 remaining subjects.

Data Analysis

Paired t tests were used to check for significant differences between mean scores obtained by direct observation versus mean scores derived from parent interview. Separate paired t tests were conducted for items, subscale scores, and the mean Motor and Cognitive domain scores. The ICC was computed to determine agreement among ratings. The ICC is a reliability coefficient calculated from estimates obtained through analysis of variance and specifically examines agreement among raters (Portney & Watkins, 1993). The ICC model (2,1) (Shrout & Fleiss, 1979) was computed in the current investigation because it includes variance components related to method of administration and provides a conservative estimate of agreement (Portney & Watkins, 1993).

The relation of scores between the two administra-
The ICC values for individual WeeFIM items ranged from .41 (Social Interaction) to .98 (Walk/Wheelchair/Crawl). The subscale ICC values ranged from .66 (Social Cognition) to .94 (Locomotion). The ICC value for total WeeFIM ratings was .93. The results of the paired t tests revealed no significant differences between scores obtained by the two administration methods for WeeFIM items, subscales, domains, or total ratings (see Table 1).

The results indicate that the Locomotion subscale (ICC = .94) demonstrated the highest consistency and the Social Cognition subscale (ICC = .66) the lowest. The low ICC values associated with items in the Cognitive domain (Social Interaction, Problem Solving, Memory) and some Self-Care items (e.g., Grooming) may reflect low variability within ratings recorded in these areas. For instance, for the Social Interaction item, 87% of the ratings were from 1 to 3, 11% were from 4 to 5, and 2% were a 6. There were no ratings of 7 for this item. Similarly, for the Grooming item, 88% of the ratings were from 1 to 3, 11% were either a 4 or 5, and there were none for 6 or 7. These two items demonstrate a restricted range of scores and are associated with low ICC values (.41 for Social Interaction, .50 for Grooming). In contrast, the Bowel Management item had an ICC value of .79 and exhibited a much broader range of ratings: 52% of the scores were 1 or 2; 25% were a 3, 4, or 5; and 23% were a 6 or 7.

In discussing the problem of restriction in the range of ratings and the impact on reliability and agreement indexes, Portney and Watkins (1993) stated that low variability among scores “will tend to decrease the correlation coefficient or any reliability coefficient” (p. 525). They suggested computing the method error coefficient of variation to address the problem of low variability. Portney and Watkins further noted, “Method error will not be affected by a restriction in range, because it looks only at the difference scores. Therefore, in situations where reliability coefficients are misleading, method error provides a useful alternative” (p. 525). For example, the method error coefficient of variation for the Walk/Wheelchair/Crawl item was 5.2%, indicating that the variation in measurement for this item across the two methods of administration was 5.2%. Because the method error is based on the variability within difference scores, it does not account for systematic variation between the set of two ratings (observation vs. interview). Portney and Watkins suggested that a paired t test be performed between the method error coefficient of variation scores. The t ratio is obtained by dividing the mean of the difference scores by the standard error of the difference scores. For the Walk/Wheelchair/Crawl item, the paired t-test value was .86 (p > .05), indicating no significant difference in the variability across the two methods of administration. The method error coefficient of variation and accompanying paired t test was performed for all WeeFIM items, subscales, domains, and total WeeFIM ratings. No significant values were found. The method error analysis suggests that no significant degree of variability existed across the two methods of administration when reduced variability was controlled. Further, the relationship between the two methods of administration was graphically examined by computing scatter plots for the Motor and Cognitive domain scores and total WeeFIM scores, which demonstrate a linear relationship among the data (see Figures 2 and 3).

Discussion

The results suggest that the two methods of WeeFIM administration (observation and interview) produce ratings that are very similar. The ICC for total WeeFIM scores was .93. The agreement among the 15 items comprising the WeeFIM Motor domain was higher (ICC = .93) than that for the Cognitive domain (ICC = .75). The lower agreement for the Cognitive domain items was believed to be due, in part, to the reduced variability of responses obtained for these items. Many of the subjects received ratings that clustered at the low end of the scale (from 1 to 3) on several of the items in the Cognitive domain. Items in the Cognitive domain (e.g., Problem Solving) demand complex conceptual skills. To successfully complete items at the higher levels (ratings of 6 and 7) requires making reasonable, safe, and timely decisions and initiating, sequencing, and self-correcting various activities (Guide for the Functional Independence Measure for Children, 1993). Young children (ages 1–4 years) cannot be expected to achieve functional independence (level 6 or 7) in items such as Problem Solving.

Those Cognitive domain items with restricted variability were associated with lower ICC values. Lower ICC values do not necessarily indicate a lack of consensus across the ratings obtained by different observation methods. This was reflected in the fact that there were no significant differences in the paired t tests for ratings obtained by both methods (see Table 1). In addition, the method error coefficient of variation analysis suggested
Table 1
Means, Standard Deviations, ICC, and Paired t Test Values for WeeFIM®

<table>
<thead>
<tr>
<th>WeeFIM Variable</th>
<th>Observation M (SD)</th>
<th>Interview M (SD)</th>
<th>ICC</th>
<th>Paired t Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Care</td>
<td>16.7 (5.8)</td>
<td>17.1 (5.1)</td>
<td>.87</td>
<td>0.95</td>
</tr>
<tr>
<td>Eating</td>
<td>5.8 (1.4)</td>
<td>5.6 (1.6)</td>
<td>.83</td>
<td>1.23</td>
</tr>
<tr>
<td>Grooming</td>
<td>2.1 (1.1)</td>
<td>2.2 (0.8)</td>
<td>.50</td>
<td>0.55</td>
</tr>
<tr>
<td>Bathing</td>
<td>2.0 (1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dressing—Upper Body</td>
<td>2.6 (0.85)</td>
<td>2.6 (0.93)</td>
<td>.79</td>
<td>0.63</td>
</tr>
<tr>
<td>Dressing—Lower Body</td>
<td>2.5 (1.1)</td>
<td>2.5 (0.97)</td>
<td>.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Toileting</td>
<td>2.6 (2.4)</td>
<td>2.3 (1.6)</td>
<td>.69</td>
<td>1.15</td>
</tr>
<tr>
<td>Sphincter Control</td>
<td>6.1 (5.2)</td>
<td>5.8 (3.7)</td>
<td>.81</td>
<td>0.96</td>
</tr>
<tr>
<td>Bladder Management</td>
<td>3.1 (2.6)</td>
<td>2.6 (1.8)</td>
<td>.76</td>
<td>1.81</td>
</tr>
<tr>
<td>Bowel Management</td>
<td>3.5 (2.6)</td>
<td>3.2 (2.1)</td>
<td>.75</td>
<td>0.92</td>
</tr>
<tr>
<td>Transfers</td>
<td>10.1 (4.7)</td>
<td>11.5 (6.7)</td>
<td>.75</td>
<td>1.11</td>
</tr>
<tr>
<td>Chair/Wheelchair</td>
<td>5.3 (2.6)</td>
<td>5.4 (2.5)</td>
<td>.93</td>
<td>0.57</td>
</tr>
<tr>
<td>Toilet</td>
<td>3.7 (2.7)</td>
<td>3.9 (2.9)</td>
<td>.58</td>
<td>0.49</td>
</tr>
<tr>
<td>Tub/Shower</td>
<td>2.6 (2.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locomotion</td>
<td>9.9 (4.3)</td>
<td>10.2 (4.5)</td>
<td>.94</td>
<td>0.28</td>
</tr>
<tr>
<td>Walk/Wheelchair/Crawl</td>
<td>5.7 (2.3)</td>
<td>5.6 (2.3)</td>
<td>.98</td>
<td>0.44</td>
</tr>
<tr>
<td>Stairs</td>
<td>4.3 (2.2)</td>
<td>4.6 (2.2)</td>
<td>.80</td>
<td>1.31</td>
</tr>
<tr>
<td>Communication</td>
<td>6.4 (2.6)</td>
<td>6.8 (2.3)</td>
<td>.71</td>
<td>1.02</td>
</tr>
<tr>
<td>Comprehension</td>
<td>2.1 (1.1)</td>
<td>3.6 (1.4)</td>
<td>.53</td>
<td>1.51</td>
</tr>
<tr>
<td>Expression</td>
<td>3.0 (1.6)</td>
<td>3.2 (1.2)</td>
<td>.68</td>
<td>0.88</td>
</tr>
<tr>
<td>Social Cognition</td>
<td>8.2 (3.1)</td>
<td>8.8 (2.8)</td>
<td>.66</td>
<td>0.73</td>
</tr>
<tr>
<td>Social Interaction</td>
<td>2.4 (1.2)</td>
<td>2.7 (1.1)</td>
<td>.41</td>
<td>0.75</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>2.5 (0.90)</td>
<td>2.7 (1.1)</td>
<td>.54</td>
<td>1.03</td>
</tr>
<tr>
<td>Memory</td>
<td>3.2 (1.3)</td>
<td>3.4 (1.0)</td>
<td>.57</td>
<td>0.95</td>
</tr>
<tr>
<td>WeeFIM Domains</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td>43.2 (17)</td>
<td>45.0 (17)</td>
<td>.93</td>
<td>1.12</td>
</tr>
<tr>
<td>Cognitive</td>
<td>13.9 (5.5)</td>
<td>15.2 (4.8)</td>
<td>.75</td>
<td>1.03</td>
</tr>
<tr>
<td>Total WeeFIM</td>
<td>57.9 (20.1)</td>
<td>60.57 (18.6)</td>
<td>.93</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Note. ICC = intraclass correlation coefficient; WeeFIM® = Functional Independence Measure for Children.

a N = 30

b Bathing and Tub/Shower transfers not observed.

c Reliability computed without Bathing and Tub/Shower transfer scores.

that the amount of variability across the two methods of administration was stable (see Figures 2 and 3).

The WeeFIM is most often used as an interview assessment to obtain information from the parent or teacher about a child’s “usual” functional performance; therefore, the person being interviewed needs to know how the child performs functional skills on a daily basis. In this study, the examiner was unknown to the subjects. During the observational administration of the WeeFIM, activities that involved physical skills (e.g., Locomotion, Transfers, Self-Care) were easier to observe than abstract skills (e.g., Comprehension, Memory, Problem Solving). The abstract nature of the Cognitive and Social Skill items may have contributed to the lower reliability values for items in the Cognitive domain. The mean ratings for the Cognitive and Communication items tended to be higher for the interview administration (see Table 1). Parents may have rated their children higher in these areas because they had more exposure to the subject’s ability than the examiner.

Figure 2. Scatter plot of total Functional Independence Measure for Children (WeeFIM®) ratings obtained by observation and parental interview.
could obtain in a 1-hour or 2-hour observation period.

Expectations across environments may have differed and affected the scoring criteria used for direct observation versus parent interview. Because independence in daily living skills is often a goal of educational and therapy programs, many subjects were provided the time to dress, undress, and wash independently in the school environment. At home, these tasks might be completed for the child by the caregiver in order to save time. The impact of environment and parental versus teacher expectations on the functional performance of children with disabilities is an area that requires additional research, both with children and their parents or teachers.

Because the WeeFIM is a new assessment, further research examining validity, reliability, and sensitivity is needed. A replication of this study using both administration methods in the same setting (i.e., observation in school) with teacher interviews might control for some of the environmental and role expectation differences referred to previously. The impact of training and influence of different professional backgrounds (e.g., educators vs. therapists) on administering, rating, and interpreting the WeeFIM also require empirical study. There may, for example, be variations in how different professional disciplines observe and score selected items. Future research on these questions will provide the information necessary to establish functional evaluation as a regular part of the measurement process for children with disabilities and members of their families.

Acknowledgment
This research was partially supported by a grant from the U.S. Department of Health and Human Services, Bureau of Maternal and Child Health (grant #MCJ-360646-010). The study was completed in partial fulfillment of the first author’s requirements for Master of Science in occupational therapy, State University of New York at Buffalo.

References


