A Quantitative Approach to Movement Skill Assessment for Children With Mental Retardation

Joonkoo Yun
Oregon State University

Deborah R. Shapiro
Georgia State University

This study examined the psychometric properties of Ulrich’s (1988) Actual Physical Competence Scale for children with mental retardation. A total of 139 children with MR, 7 to 13 years of age participated. Confirmatory factor analyses indicated that a multidimensional model of skill assessment captures the motor performance of those with MR more accurately than a unidimensional model. Indices of goodness of fit for the multidimensional model were GFI = .91, RMSEA = .09, (χ²/df) = 2.15, and CFI = .93. Test-retest reliability and internal consistency for the total test battery was ICC = .91 and α = .62, respectively. When evaluating movement skills of children with mental retardation, a multidimensional model incorporating both locomotor and object control skills is recommended.

The past 25 years have brought considerable agreement among researchers and practitioners on the interrelationships and differences between the concepts of movement skills, motor abilities, and general motor abilities (Burton & Rodgerson, 2001). Traditional taxonomies place movement skills at the highest level of the theoretical hierarchy followed by motor abilities and general motor abilities at the base. Movement skills are defined as a group of movements that have a similar movement form or pattern and similar function such as running or throwing. Motor abilities are defined as general traits or capacities that underlie performance of a variety of movement skills (Fleishman, 1972). General motor ability (GMA) refers to a single trait of an individual that underlies performance of all movement skills (Burton & Rodgerson, 2001).

Burton and Rodgerson (2001) present several concerns with the hierarchical taxonomy of movement skills and propose a reconceptualization of the constructs. Their new taxonomy is purported to (a) assist with interpretation of research, (b) aide in consistency of use of terms and constructs, (c) recognize increased specificity of movement skills across childhood outlined in current research, and (d) be more...
congruent with current assessment practices in adapted physical education. The new taxonomy consists of four levels: movement skills, movement skill sets, movement skill foundations, and general motor ability. This paper will address only the components of movement skills and movement skill sets. Movement skills in this paper are similarly defined as in the old taxonomy. Movement skills are differentiated by the limbs and joints used and the timing, force, and sequence characteristics of the motions (Burton & Rodgerson, 2001). An important factor in the assessment of movement skills is the range of contexts and task requirements under which the skills can be performed. Movement skills that cluster together into assessment instrument subsets or divisions are considered to be movement skill sets, with skills grouped according to criteria related to categories or classifications (Burton & Rodgerson, 2001). Movement skill sets are typically determined by factor analysis of large groups of skills resulting in either uni- or multidimensional skill sets that are justified by their relationship to the theoretical perspective on which they were analyzed. The locomotor and object control subdomains in the Test of Gross Motor Development-2 (Ulrich, 2000) and the five subsets in the Movement Assessment Battery for Children Checklist (Henderson & Sugden, 1992) are examples of multidimensional movement skill subsets indicating the function of skill sets, the type of muscle coordination and control needed by the performer, or performer attributes (Burton & Rodgerson, 2001).

Inclusion of students with mental retardation (MR) in general physical education classes has highlighted for educators and researchers alike both qualitative and quantitative differences in movement skill performance between students with and without MR. Not only is there lower level and rate of learning for movement skills, but there are qualitative differences in movement skill performance as well (DiRocco, Clark, & Phillips, 1987; Wade, 1990). Additionally, children with MR may experience neurological differences that limit their movement skill development (Block, 1991; Maraj, Hillman, Jeannonne, & Ringenbach, 2003; Sparrow & Day, 2002). Such neurological differences may influence the rate and order of movement skill acquisition in children with MR (Block 1991; Sugden & Keogh, 1990; Vermeer, 1995).

The qualitative differences in movement skills of individuals with and without MR may reflect different movement strategies and patterns of behavior by which to successfully achieve the task goal (Newell, 1986; Keogh & Sugden, 1985). A qualitative performance assessment highlights potential strategy differences used in skill performance with little regard to movement outcome. Burton and Miller (1998) argue that movement outcome is an inherent part of functional movement skills. Functional movement skills should ultimately be described by the outcome of movement behavior rather than through descriptions of movement encompassing kinematic characteristics.

To evaluate movement skill performance in individuals with MR in both clinical and physical education settings, two types of standardized assessment instruments are commonly used. These instruments measure movement either quality/process or quantity/product. Instruments measuring movement process evaluate the pattern of movement displayed by a child and provide qualitative information on performance. The use of process oriented assessments have been encouraged extensively in physical education setting (e.g., Anderson & Goode, 1997; Gallahue & Ozman, 2002) as they (a) identify children who are significantly below age norms; (b) enable teachers to assess the respective movement sequence
of each child individually; (c) provide a basis for program planning; (d) verify the
effectiveness of instructional strategies, goals, and objectives through observation;
and (e) assess changes in motor performance as a function of age or experience
(Haubenstricker & Seefeldt, 1986).

Most standardized quantitative oriented assessment tools such as the
Bruininks-Oseretsky Test of Motor Proficiency (BOTMP; Bruininks, 1978) are
intended to evaluate motor ability (Burton & Miller, 1998). Motor ability items
often include balance, agility, reaction time, and manual dexterity. Although the
methods of measuring these variables are product oriented, the intent of these
instruments is to measure the underlying construct, motor ability, to predict future
performance (Davis, 1984; Newell, 1977) and make normative comparisons to
identify performance levels within a population. For example, BOTMP measures a
child’s response speed by dropping a stick and measuring the distance the stick drops
before it is stopped. BOTMP is measuring response time (outcome of movement
behavior) under the assumption that response time is an underlying cause of future
movement performance.

Ulrich’s (1988) Actual Physical Competence Scale (APCS) employs a
quantitative approach to capture the level of movement skills within a given
constraint rather than measuring motor ability to predict future performance. All
APCS items are physical activities that related to elementary physical education
classes and attempt to capture the actual movement skills of children with MR,
ages 7 to 13 years, using an outcome based assessment approach. For example,
actual movement skill at jumping rope was measured by the number of successful
jumps out of 5 attempts rather than measuring underlying constructs of jumping
such as balance or coordination. Using the APCS, Yun and Ulrich (1997) studied the
theoretical relationship between perceptions of movement competence with actual
movement skill performance of children with MR. The construct of actual movement
competence as studied by Yun and Ulrich is synonymous with the phrase movement
skill set used throughout this paper when describing the clustering of movement
skills. Although the APCS appeared promising as a way of quantitatively measuring
the nature of movement skills, its psychometric properties have not been examined.
Using the first two components of Burton and Rodgerson’s (2001) taxonomy, the
purpose of the present study was to examine the psychometric properties of Ulrich’s
(1988) APCS for assessing quantitative movement skill performance of children
with MR. Therefore, is evidence of validity and reliability for the APCS stronger
for unidimensional or multidimensional movement skill set clusters?

Method

Participants

A total of 139 children (69 males and 70 females) from both urban and rural school
districts ranging in age from 7 to 13 years ($M = 9.61$, $SD = 1.70$) participated in
this study. All participants exhibited mild MR as reported by their special education
teacher, had IEPs on file at their schools, and received special education services.
Exact information regarding IQ scores was not available. School administrators
confirmed that all children met the state guidelines of mild MR: (a) performed
between two and three standard deviations below the mean on an intelligence test
and (b) exhibited an adaptive behavior profile within the range of mild mental
retardation. None of the participants had any known physical constraints that could influence performance. Among 139 children, data from 109 children were from a previously published study by Yun and Ulrich (1997). Additional data were collected for test-retest reliability estimation. An independent $t$ test and Levene’s test for equality of variances were employed to examine possible bias between the previous and additional groups. The analysis suggested that the two groups have equal variance, $F(1, 137) = 2.15, p = .16$ and that there is no group difference on motor performance, $t = -.46, p = .65$. Based on these analyses, the two samples were assumed to represent the same population.

Instrumentation and Data Collection Procedures

APCS evaluates motor skill performance in a variety of object control and locomotor areas including (a) dribbling a ball in and out of a series of cones, (b) shooting a basketball from a distance of 10 feet, (c) jumping rope, (d) throwing a ball to a target, (e) catching a tennis ball, (f) batting a tossed ball, (g) kicking a stationary ball to a target, (h) running 40 yards, (i) skipping for 30 seconds, and (j) performing a standing long jump (see Yun & Ulrich, 1997 for additional information). Table 1 illustrates each APCS item. The selected skills represent those commonly performed by children with and without MR in physical education and/or in community sport programs.

We tested participants individually on each of the 10 skills in a quiet area. Motor skill assessments took approximately 30 minutes to complete during the participant’s regularly scheduled physical education class. Prior to performing each skill, verbal instruction with a demonstration was performed and one practice trial was permitted. They are incorporated into the standardized procedures for this assessment. Upon completion of the test items, participants returned to their class. A total of 24 participants from the second data collection group completed the APCS a second time. Results of this second testing session were used to provide evidence of test-retest reliability.

Data Analysis

Data analysis for this study consisted of examining the internal structure of APCS by employing a confirmatory factor analysis using AMOS 4.0 software and investigating reliability by employing test-retest and internal consistency. We specified two models to examine the structure of APCS. The first model, the null model, is based on the instrument’s original use as described by Yun and Ulrich (1997). That is, movement skills on the APCS represent a unidimensional movement skill set or construct employing a total motor performance test score. In the null model, all items were entered on a single construct, named actual competence. The second model was an alternative model, which was based on the notion that movement skills can be clustered into two sets generally referred to as locomotor and object control subdomains (Burton & Miller, 1998; Gallahue & Ozmun, 2002; Payne & Isaacs, 2002). Burton and Rodgerson (2001) recommended that individual skills and skill sets within an assessment instrument be labeled and organized according to function rather than the attributes or abilities needed for skill performance. In accordance with their recommendation and consistent with a multidimensional framework, we divided the skills into two groups. Standing
<table>
<thead>
<tr>
<th>Items</th>
<th>Description of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long Jump</td>
<td>The participant starts with both feet positioned behind a line. The score will be the distance jumped calculated to the nearest 1/2 inch.</td>
</tr>
<tr>
<td>Running</td>
<td>The participant runs a shuttle run a distance of 90 feet. Two cones are placed so their outer edges are 45 feet apart. The score will be recorded to the nearest 1/10th of a second.</td>
</tr>
<tr>
<td>Batting</td>
<td>The participant strikes a 4 inch plastic ball using a light-weight plastic bat. The ball is tossed slowly by the examiner using an underhand delivery a distance of 15 feet. The score will be the number of successful attempts out of five attempts where the participant makes contact with the ball and the ball travels in a forward direction.</td>
</tr>
<tr>
<td>Shooting</td>
<td>The participant shoots a mini-size basketball into a basketball hoop, positioned at regular height. The participant is positioned five feet in front of the rim of the hoop. The score will be the number of successful attempts out of five attempts.</td>
</tr>
<tr>
<td>Kicking</td>
<td>The participant kicks a stationary 7 1/2 inch sponge ball into a goal from a distance of 18 feet. The goal is five feet wide by five feet high. The score will be the number of successful attempts out of five attempts.</td>
</tr>
<tr>
<td>Throwing</td>
<td>The participant throws a tennis ball at a 24 inch square target from a distance of 10 feet. The score will be the number of successful attempts out of five attempts.</td>
</tr>
<tr>
<td>Skipping</td>
<td>The participant skips for 30 seconds. The score will be the number of completed step hop cycles performed. One cycle is a step and a hop on the right and left foot.</td>
</tr>
<tr>
<td>Catching</td>
<td>The participant catches a tennis ball. The ball is tossed slowly using an underhand delivery from a distance of 10 feet. The score will be the number of successful attempts out of five attempts.</td>
</tr>
<tr>
<td>Jumping Rope</td>
<td>The participant jumps a rope that is turned by the participant. The score will be the number of successful attempts out of five attempts to jump the rope.</td>
</tr>
<tr>
<td>Dribbling</td>
<td>The participant dribbled a ten inch playground ball for 30 seconds. Six cones are placed in a straight line parallel to a wall. The cones should be 5 feet form the wall and other cones. The participant is required to dribble the ball around the cones. The score will be the number of cones passed in 30 seconds.</td>
</tr>
</tbody>
</table>

*Note.* Based on original Actual Physical Competence Scale, which was printed in Yun & Ulrich (1997).
long jump, running, skipping, and jumping rope comprised the locomotor skills construct. Batting, throwing, catching, kicking, shooting, dribbling, and jumping rope were grouped under the object control skills construct. Because jumping rope includes both locomotor and object control skills, this item was entered on both constructs.

As there is no universally accepted goodness of model fit index, the null and alternative models were evaluated using four common indices: (a) the goodness of fit index (GFI), (b) the root mean square error of approximation (RMSEA), (c) the ratio of chi-square and degree of freedom ($\chi^2/df$), and (d) the comparative index (CFI). GFI, developed by Jöreskog and Sörbom (1984, cited in Arbuckle & Wothke, 1995), is considered to be better than any other absolute fit index (Marsh, Balla, & McDonald, 1988). CFI was examined as it tends to be a more consistent estimation in samples smaller that 250 (Hu & Bentler, 1995). As a general guideline, cutoff scores for GFI and CFI are .90 and .95, respectively. The ratio of chi-square and degrees of freedom ($\chi^2/df$) is another indicator of test structure fitness and a ratio of 3 or less is acceptable (Carmines & McIver, 1981). Browne and Cudeck (1993) suggested that RESEA value of .05 or less is considered to be an excellent fit and .10 or greater as a poor fit.

Internal consistency was determined using Cronbach $\alpha$ on the total test as well as both the locomotor and object control skills subtests. The standard error of measurement (SEM) also was estimated. Test-retest reliability was computed using intraclass correlation. Test-retest reliability was conducted on 24 participants within a period of 2 weeks between the first and second test session. Test-retest reliability was determined for both the total test as well as for each individual item.

**Results**

**Null Model**

The Critical Ratios (CR) allows us to examine the statistical link between the construct and each item. CR is equivalent to the conventional $t$ statistics (Arbuckle & Wothke, 1995). CR from the confirmatory factor analysis ranged from -8.38 to 9.37 indicating all items are significantly linked with actual competence at $p < .01$ level. The standardized regression weights and squared multiple correlation from the null model are illustrated in Figure 1. The squared multiple correlations ranged from .09 to .61. Indices of goodness of fit were GFI, RMSEA, $\chi^2/df$, and CFI can be found in Table 2.

**Alternative Model**

The CRs for the alternative model ranged from -8.83 to 9.51. All items entered on the locomotor and object control skills constructs were significant ($p < .01$) with the exception of jumping rope, C.R. = -.89, $p = .38$. While jumping rope was entered on both locomotor and object control constructs in the initial model, the results indicated that jumping rope is more aligned with locomotor activity. Standardized regression weights and squared multiple correlations of the alternative model are found in Figure 2. All indices of fit for the alternative model can be found in Table 2.
Figure 1 — Null model.

Note. Estimated standardized regression weights are denoted next to the respective lines. Squared multiple correlations show percentages of variance explained by actual competence. Squared multiple correlations are denoted next to the respective items. Significant effects are indicated by solid arrows.
Reliability Estimation

Test-retest reliability results were determined using ANOVA. Test-retest reliability for the total test battery was ICC = .91. Locomotor subtest and object control subtest were ICC = .89 and .85, respectively. Reliability and descriptive statistics for each skill can be found in Table 3. The test-retest reliability for the total test battery, locomotor and object control subtests, and standing long jump were considered to be good; they were greater than .80. The reliability for running, shooting a basketball, and skipping were satisfactory (> ICC = .70), while the reliability for batting and catching were acceptable (> ICC = .60). Item reliability for the throwing, dribbling, jumping rope, and kicking were questionable. Internal consistency is a reliability measure that examines how well the individual items measure the unidimensional construct of actual physical competence. Internal consistency reliability obtained from all 10 activities for the APCS was \( \alpha = .62 \). When internal consistency reliability coefficients were calculated from locomotor and object control skills, respectively, Cronbach alpha dropped to \( \alpha = .32 \) (4 items) and \( \alpha = .48 \) (7 items). SEM of the total battery was .15. SEM for each item ranged from .56 to 4.91.

Discussion

In the null model, goodness of fit indices met adequate levels with the exception of the chi-square and degrees of freedom ratio, \( \chi^2/df = 2.56 \). This finding suggests that the movement skills on the APCS should not be viewed as a unidimensional construct or a single movement skill set. In the alternative model, goodness of fit indices considerably improved. All indices except CFI indicated that the model was a fit at a satisfactory level. Although CFI did not reach the stringent recommended cutoff score of greater than .95 (Hu & Bentler, 1995), it substantially improved from .89 to .93. A score of .90 is considered minimally acceptable (Kline, 1998). Considering each fit index reflects only a particular aspect of fit, all four different indices met at least acceptable levels strengthening the validity evidence for the alternative, multidimensional model. Yun and Ulrich (2002) argued that construct related validity becomes stronger when empirical tests and theoretical models are congruent. The results of the improvements on goodness of fit indices demonstrate the multidimensional nature of movement skills sets in the APCS (Burton & Miller, 1998; Gallahue & Ozmun, 2002; Payne & Isaacs, 2002).

<table>
<thead>
<tr>
<th>Model</th>
<th>GFI</th>
<th>RMSEA</th>
<th>( \chi^2/df )</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model</td>
<td>.89</td>
<td>.11</td>
<td>2.56*</td>
<td>.89</td>
</tr>
<tr>
<td>Alternative Model</td>
<td>.91*</td>
<td>.09*</td>
<td>2.15*</td>
<td>.93*</td>
</tr>
</tbody>
</table>

Note: * Meet a general guidelines for manual level of cutoff score.
Figure 2 — Alternative model.

Note: Squared multiple correlations show percentages of variance explained by Locomotor or Object control skills. Squared multiple correlations are denoted next to the respective items. Significant effects are indicated by solid arrows.
The magnitude of the test-retest reliability coefficient for the total battery was considered to be good (> .80), and internal consistency exceeded the minimal requirement of .60 (Salvia & Ysseldyke, 1988). Test-retest reliability of both locomotor and object control subtests were also greater than .80. However, internal consistency is troublesome for the individual subtests due largely to having too few items in each subscale. This is a reflection of the APCS not being designed to assess a multidimensional construct of motor skill performance. Since the number of items heavily influences internal consistency (Crocker & Algina, 1986), the appropriate number of items per subscale to increase reliability was calculated using Spearman-Brown prophecy formula. A total of 13 locomotor and 12 object control skills would be required to meet minimal levels of internal consistency.

A few items were found to demonstrate low test-retest reliability. It is interesting to note that items measuring accuracy of performance (kicking and throwing activities) demonstrated lower test-retest reliability. These items should be modified with regard to changing task constrains or changing the scoring system to improve reliability. Until such changes are made, items with low reliability should not be eliminated because of issues related to content-relevance. APCS was developed to capture actual overall movement competences of children. Eliminating an activity that represents a common physical activity for young children can limit the instrument’s content relevance. Considering results from all test-retest reliability and internal consistency evidence, use of the total test battery is required to maintain an acceptable level of reliability.

### Table 3  Descriptive Statistics and the Results of Item Test-Retest Reliability

<table>
<thead>
<tr>
<th>Skill</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Reliability Coefficient</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standing long jump</td>
<td>41.25</td>
<td>12.64</td>
<td>10 – 72</td>
<td>.92</td>
<td>3.62</td>
</tr>
<tr>
<td>Running (sec.)</td>
<td>9.87</td>
<td>1.75</td>
<td>3.7 – 18.9</td>
<td>.74</td>
<td>0.56</td>
</tr>
<tr>
<td>Batting a tossed ball</td>
<td>2.94</td>
<td>1.58</td>
<td>0 – 5</td>
<td>.61</td>
<td>1.04</td>
</tr>
<tr>
<td>Shooting a basketball</td>
<td>1.10</td>
<td>1.26</td>
<td>0 – 5</td>
<td>.73</td>
<td>0.66</td>
</tr>
<tr>
<td>Kicking a stationary ball</td>
<td>2.29</td>
<td>1.28</td>
<td>0 – 5</td>
<td>.08</td>
<td>1.11</td>
</tr>
<tr>
<td>Throwing a ball</td>
<td>3.68</td>
<td>1.30</td>
<td>0 – 5</td>
<td>.26</td>
<td>1.06</td>
</tr>
<tr>
<td>Skipping</td>
<td>18.83</td>
<td>9.30</td>
<td>0 – 31</td>
<td>.75</td>
<td>4.91</td>
</tr>
<tr>
<td>Catching a ball</td>
<td>3.90</td>
<td>1.31</td>
<td>0 – 5</td>
<td>.66</td>
<td>0.79</td>
</tr>
<tr>
<td>Jumping rope</td>
<td>3.14</td>
<td>2.02</td>
<td>0 – 15</td>
<td>.59</td>
<td>1.53</td>
</tr>
<tr>
<td>Dribbling a ball</td>
<td>17.90</td>
<td>8.49</td>
<td>3 - 28</td>
<td>.58</td>
<td>3.81</td>
</tr>
<tr>
<td>Locomotor skills</td>
<td></td>
<td></td>
<td></td>
<td>.89</td>
<td>.43</td>
</tr>
<tr>
<td>Object control skills</td>
<td></td>
<td></td>
<td></td>
<td>.85</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>.91</strong></td>
<td><strong>.15</strong></td>
</tr>
</tbody>
</table>

*Note.* Sample size for descriptive statistics was 139. Reliability and SEM was estimated from 24 participants.
Assessing Movement Skills

Summary

The results of the present study support the taxonomy of Burton and Rodgerson (2001) in which they proposed that the grouping of movement skills reflect similarity in function rather than attributes or abilities needed to perform the skills. The unidimensional grouping of movement skills on the APCS held the minimal levels of reliability, with questionable validity evidence. The skills assessed formed movement skill sets that fit a multidimensional model operationally defined as object control and locomotor skills. The alternative multidimensional model had stronger validity evidence than the original model, with additional items needed to improve reliability. It is recommended that when evaluating movement skill sets of children with MR that researchers and practitioners use a multidimensional model incorporating both locomotor and object control skill constructs.

References


Acknowledgment

The authors wish to thank Dr. Dale A. Ulrich at the University of Michigan for allowing the use of his data as a part of this study.