Physical Activity Measurement Among Individuals With Disabilities: A Literature Review

Carlos M. Cervantes
Texas State University

David L. Porretta
The Ohio State University

This review examined the literature on physical activity measurement among individuals with disabilities utilizing Yun and Ulrich’s (2002) view on measurement validity. Specific inclusion criteria were identified. The search produced 115 articles; however, only 28 met all specified criteria. Findings revealed that self-reports and accelerometers were the most common approaches to measuring physical activity, and individuals with orthopedic impairments, those with mental retardation, and those with other health impairments received the most attention. Of the 28 articles, 17 (61%) reported validity and reliability evidence. Among those studies reporting validity, criterion-related evidence was the most common; however, a number of methodological limitations relative to validity were observed. Given the importance of using multiple physical activity measures, only five (18%) studies reported the use of multiple measures. Findings are discussed relative to conducting future physical activity research on persons with disabilities.

Historically, the physical activity needs of individuals with disabilities have received little attention, but recently have become a national agenda (Rimmer, Braddock, & Pitetti, 1996; Rimmer & Braddock, 2002; Rimmer, Wolf, Armour, & Sinclair, 2007). These individuals demonstrate reduced levels of physical activity and higher prevalence of overweight and obesity, which increase their risk for developing secondary health conditions (Cooper et al., 1999; Durstine et al., 2000; Kinne, Patrick, & Lochner Doyle, 2004). According to Healthy People 2010 (United States Department of Health and Human Services [USDHHS], 2000), 55% of adults with disabilities do not engage in any leisure-time physical activity compared with 37% of adults without disabilities (Centers for Disease Control & Prevention [CDC], 2007). One of the Healthy People 2010 objectives is to reduce to 20% the number of individuals (regardless of disability status) who do not engage in any leisure-time
physical activity (USDHHS, 2000). Similarly, only 18% of adults with disabilities reported engaging in regular moderate to vigorous physical activity (MVPA), 30 min or more per day for 5 or more days of the week when compared with 33% of adults without disabilities (CDC, 2007).

Similar to their adult counterparts, youth with disabilities tend to be less physically active, less physically fit, and have higher prevalence of overweight/obesity than their peers without disabilities (Hogan, McLellan, & Bauman, 2000; Longmuir & Bar-Or, 2000; Whitt-Glover, O’Neill, & Stettler, 2006). Rimmer, Rowland, and Yamaki (2007), in a secondary analysis of data from the 2005 Youth Risk Behavior Surveillance System (YRBSS), indicated a greater prevalence of overweight among adolescents with disabilities (16.7%) compared with age-matched adolescents without disabilities (12.8%). The need for individuals with disabilities to be physically active is even more important based on the paucity of research on physical activity interventions (Heller, Hsieh, & Rimmer, 2004; Taylor, Baranowski, & Young, 1998) and the lack of physical activity guidelines and recommendations for this subgroup of the U.S. population (Cooper et al., 1999).

Physical activity has been widely accepted as any bodily movement resulting in energy expenditure (Caspersen, Powell, & Christenson, 1985). Under this broad definition, physical activity includes day-to-day body movements such as walking, climbing stairs, biking for transportation, house work, and gardening (Biddle, 1994). For the purpose of this paper, physical activity refers to any voluntary body movement produced by muscles resulting in energy expenditure. The term “voluntary” has been included to eliminate involuntary body movements that may characterize certain disabilities (e.g., cerebral palsy) because involuntary movements are typically not under the individual’s control. Physical activity can be quantified or categorized by type (activity), duration (how long), frequency (how often), and intensity (how hard; LaMonte, Ainsworth, & Reis, 2006; Sallis & Saelens, 2000).

With the recognition of increasing levels of recommended physical activity as a national health priority, the identification and establishment of physical activity measures that produce valid and reliable scores as well as improving the accuracy of available assessment techniques for physical activity have become major research priorities (Bauman, Phongsavan, Schoeppe, & Owen, 2006; Dishman, Washburn, & Schoeller, 2001). Validity has traditionally been viewed as the degree to which a test and/or instrument measures what it purports to measure (Thomas, Nelson, & Silverman, 2005). The meaning of validity has changed over time, however, shifting from the view of a characterize of a test and/or measure to that of an evaluative judgment of the extent to which meaningful interpretations can be made from measurement scores (Messick, 1995; Rowe & Mahar, 2006; Yun & Ulrich, 2002). In 2002, Yun and Ulrich described validation procedures pertinent to adapted physical activity, recommending the reporting of several categories of validity evidence: (a) content-related, (b) criterion-related, and (c) construct-related. These categories of validity evidence rather than separate types of validity (e.g., content validity, criterion validity) are viewed as interdependent validity sources (Yun & Ulrich, 2002). This paper focuses on Yun and Ulrich’s sources of validity evidence as it refers to adapted physical activity applications.
Estimating measurement validity is critical for research on physical activity among individuals with disabilities because it provides the foundation for interpreting scores derived from the different measures of physical activity and decision making processes (Yun & Ulrich, 2002). The use of physical activity measures that produce valid and reliable scores is necessary because it serves to (a) identify current levels of physical activity among different groups in the population, (b) document the frequency and distribution of physical activity among particular groups, (c) examine the dose-response (i.e., how much physical activity) relationship between physical activity and health, (d) identify correlates of physical activity, and (e) evaluate program effectiveness (Sirard & Pate, 2001; Trost, 2007).

Equally important are the development and refinement of physical activity measures to quantify physical activity among individuals with disabilities (Cooper et al., 1999; Rimmer et al., 1996; Warms, 2006). A panel of experts in health and disability issues has identified several areas in need of further research related to physical activity and health among individuals with disabilities; these include (a) characterizing physical activity patterns, (b) explaining physical activity impact, and (c) determining the dose-response relationship between physical activity and health outcomes (Cooper et al., 1999). These and other questions can be difficult to answer without first establishing adequate procedures to provide validity and reliability evidence, as well as identifying physical activity measures capable of producing such evidence among individuals with disabilities (Beets, Combs et al., 2007; Motl, McAuley, Snook, & Scott, 2006; Warms, 2006).

To date, a range of methods and/or techniques have been used to assess physical activity. These include (a) self-reports, (b) pedometers and accelerometers, (c) direct observation systems, (d) heart rate monitors, and (e) indirect calorimetry and doubly labeled water (Dishman, Washburn, Schoeller, 2001; Welk, 2002). Measuring physical activity behavior is a complex task (Caspersen et al., 1985; LaMonte, Ainsworth, & Reis, 2006), however, and no single measure of physical activity is capable of capturing the entire range of dimensions and/or behaviors that encompass physical activity (Dishman et al., 2001; Laporte, Montoye, & Caspersen, 1985).

A number of reviews on the existing methods to assess physical activity have been published (e.g., Dishman et al., 2001); however, only one publication to date (Warms, 2006) has focused on measurement-related physical activity issues and applications among individuals with disabilities. In Warms’ (2006) review, the strengths, weaknesses, and special considerations of existing physical activity measurement techniques for individuals with disabilities were identified. Warms also used the term disability in a broad sense and did not identify measures for validity evidence. As a result, further examination of issues related to physical activity measurement among individuals with disabilities is warranted. Describing in detail these measures, their adequacy and usefulness among individuals with disabilities is not the main purpose of this paper. For a discussion on when each physical activity measure is most appropriate, the reader is referred to Fittipaldi-Wert and Brock (2006). Therefore, using a known framework for identifying individuals with disabilities, the purpose of this review is to extend Warms’ (2006) work by reviewing the various methods of physical activity measurement with focus on those methods in which validity and reliability evidence have been reported.
Method

Individuals With Disabilities

Individuals with disabilities in this review are classified as those identified in Public Law 108–446 Individuals with Disabilities Educational Improvement Act (IDEIA) of 2004: (a) autism, (b) deaf-blindness, (c) hearing impairments, (d) mental retardation, (e) orthopedic impairments, (f) speech or language impairments, (g) visual impairments including blindness, (h) emotional disturbance, (i) learning disabilities, (j) multiple disabilities, (k) other health impairments, (l) traumatic brain injuries, and (m) deafness.

Search Strategy

Primary searches were conducted through the search engines PubMed and SPORT-Discus. The searches were designed to identify studies in which physical activity among individuals with disabilities was measured. The keywords used to identify articles were autism, developmental disability, pervasive developmental disorder, deaf-blindness, hearing impairments, mental retardation, intellectual disabilities, cognitive disabilities, orthopedic impairments, physical disability, speech or language impairments, visual impairments, blindness, emotional disturbances, learning disabilities, multiple disabilities, severe disabilities, chronic health disability and impairment, traumatic brain injuries, spinal cord injuries, and deafness. Each of the keywords was combined with physical activity, physical activity measurement, physical activity assessment, and exercise to identify the literature on the topic of interest for this review. This search produced a total of 115 articles. Secondary searches were conducted by examining the reference sections of retrieved papers to identify additional studies that may not have been identified through either of the search engines.

Inclusion and Exclusion Criteria

Inclusion criteria included (a) English language articles whose primary and/or one of the main purposes was to measure physical activity among individuals with disabilities, (b) studies in which a measure of physical activity was validated for a particular group of individuals with disabilities, (c) studies in which the target was individuals with a disabling condition identified among one of IDEIA’s disability categories, (d) studies in which physical activity data were reported, and (e) studies published from January 1990 to December 2007. Based on these criteria, 28 out of 115 articles met all inclusion criteria and were reviewed.

The review focused on articles published in peer-reviewed journals, primarily because of the accessibility to retrieve such literature. Unpublished papers from proceedings were excluded because of convenience. Inclusion criteria were applied to all retrieved papers. To determine eligibility, each article was independently reviewed by the first author and two other reviewers. If any disagreement was observed among the reviewers, the paper(s) was discussed as a group until agreement was reached. Any paper that failed to meet all inclusion criteria was
not reviewed. In addition, no studies involving the use of heart rate monitoring, indirect calorimetry, and doubly labeled water were included because they did not meet the inclusion criteria. While these are appropriate measures of energy expenditure (LaMonte, Ainsworth, & Reis, 2006), they are not direct measures of physical activity behavior. It is important to note that physical activity and energy expenditure cannot be equated, as they are two distinct constructs (LaMonte et al., 2006; Tudor-Locke & Myers, 2001).

Results

Seventeen of the reviewed studies (60.7%) reported validity and reliability evidence of the measure(s) of physical activity, while 11 studies (39.3%) reported validity coefficients only. Criterion-related validity evidence was found to be the most common approach to estimating measurement validity among the studies reviewed (22 studies, 78.6%), while the most commonly reported reliability estimates were test-retest for self-reported physical activity and interinstrument in motion sensors (see Table 1). Interinstrument reliability refers to the variability between two like-model measurement instruments (e.g., two Actical accelerometers) worn during a single trial (Esliger & Tremblay, 2006).

Self-Reported Physical Activity

Among the studies that met all inclusion criteria, eight (28.6%) used a self-report as the primary measurement source (see Table 1). These included self-administered recalls (Latimer, Martin Ginis, Craven, & Hicks, 2006; Martin Ginis, Latimer, Hicks, & Craven, 2005) and surveys (Kayes et al., 2007; Nosek, Hughes, Robinson-Whelen, Taylor, & Howland, 2006; Rimmer, Riley, & Rubin, 2001; Rimmer, Rubin, Braddock, & Hedman, 1999; van der Ploeg et al., 2007; Washburn, Zhu, McAuley, Frogley, & Figoni, 2002). Six studies were primarily designed to evaluate the validity of a self-report measure of physical activity (e.g., survey, questionnaire, recall).

Among the number of self-reports used, there were three especially designed for individuals with disabilities: (a) the Physical Activity and Disability Survey (PADS; Rimmer et al., 1999), (b) the Physical Activity Recall Assessment for People with Spinal Cord Injuries (PARA-SCI; Martin Ginis et al., 2005), and (c) the Physical Activity Scale for Individuals with Physical Disabilities (PASIPD; Washburn et al., 2002). These three measures have been further described elsewhere. The most common approaches to validity evidence reported for self-report physical activity measures were criterion-related (reported in five studies) and content and construct-related (reported in three studies each). Regarding reliability, the most common methods were test-retest and internal consistency, reported in five and four studies, respectively. The specific groups of individuals with disabilities for which evidence of validity for a self-report physical activity measure has been reported were adults with orthopedic impairments (Latimer et al., 2006; Martin Ginis et al., 2005; Nosek et al., 2006; Rimmer et al., 1999; van der Ploeg et al., 2007; Washburn et al., 2002) and adults with other health impairments (Kayes et al., 2007; Rimmer et al., 2001).
Motion Sensors: Pedometers and Accelerometers

**Pedometers.** Among the studies that met all inclusion criteria, four (14.3%) reported the use of pedometers as the primary source of measuring physical activity behavior (see Table 1). Based on the studies included in this review, validity evidence has been reported for the following groups: (a) youth with different disabilities (Beets, Combs et al., 2007), (b) youth with visual impairments (Beets, Foley, Tindall, & Lieberman, 2007), (c) adults with mental retardation (Stanish, 2004), and (d) adults with other health impairments (Motl, McAuley, Snook, & Scott, 2005). All four studies reported criterion-related validity. In three of the studies, reliability was primarily reported as interinstrument.

**Accelerometers.** Regarding the use of accelerometers among individuals with disabilities, eight studies (28.6%) met all inclusion criteria (see Table 1). Validity evidence for accelerometers has been reported among (a) adolescents and young adults with mental retardation (Kozub, 2003), (b) adults with orthopedic impairments (Bussman, Reuvenkamp, Veltink, Martens, & Stam, 1998; Warms & Belza, 2004; Washburn & Copay, 1999), (c) adults with other health impairments (Busse, Pearson, Van Deursen, & Wiles, 2004; van den Berg-Emons, Bussman, Balk, & Stam, 2000), (d) adults with traumatic brain injury (Tweedy & Trost, 2005), and (e) children and adolescents with visual impairments (Kozub, Oh, & Rider, 2005). Similar to pedometer validity, all eight studies reported criterion-related validity evidence.

**Direct Observation**

Two studies (7.2%) were reviewed in which a direct observation instrument was used as the primary source of measurement (see Table 1). It is important to note that one study specifically addressed the psychometric properties of two direct observation instruments, the System for Observing Fitness Instruction Time (SOFIT) and the Children’s Activity Rating Scale (CARS; Taylor & Yun, 2006). In both studies reporting validity evidence for the direct observation instruments, criterion-related was the preferred approach. Interobserver agreement (IOA) was used in one study as the form of reliability, while in the second study, generalizability theory was used for reliability estimation. Based on the reviewed literature, validity evidence for direct observation has been reported only among children with mental retardation (i.e., Faison-Hodge & Porretta, 2004; Taylor & Yun, 2006).

**Multiple Measures of Physical Activity**

Given the known limitations of each of the existing physical activity measures, the use of multiple measures has been suggested to provide a more comprehensive assessment of physical activity behavior (Bassett, 2000; Dishman et al., 2001). This is based on the rationale that by using multiple measures, one measure will compensate for the weakness of another. Among the studies that met all inclusion criteria, only five (17.9%) reported using multiple measures to assess physical activity (see Table 1). These included a combination of (a) self-reports, pedometers, and accelerometers (Gosney, Scott, Snook, & Motl, 2007; Motl, McAuley, Snook, & Scott, 2006); (b) accelerometers and self-reports (Pan & Frey, 2005;
Table 1  Physical Activity Studies Reporting Validity and Reliability Measures

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Measure(s)</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kayes et al., 2007</td>
<td>30 adults with Multiple Sclerosis (MS)</td>
<td>Self-report</td>
<td>Criterion (Accelerometer)</td>
<td>Test-retest: ICC = .92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*Regression analysis 95%CI</td>
<td>Bland-Altman: 95% limits of agreement (-17.4, 17.4)</td>
</tr>
<tr>
<td>Latimer et al., 2006</td>
<td>73 adults with Spinal Cord Injury (SCI)</td>
<td>Self-report</td>
<td>Convergent (Fitness measures)</td>
<td>Not applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Multiple scores</td>
<td>Multiple scores</td>
</tr>
<tr>
<td>Martin Ginis et al., 2005</td>
<td>Reliability- 102 adults with SCI; validity-14 adults with SCI</td>
<td>Self-report</td>
<td>Face, content, and criterion (Indirect calorimetry) Multiple scores</td>
<td>Test-retest</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Multiple scores</td>
</tr>
<tr>
<td>Nosek et al., 2006</td>
<td>386 women with orthopedic impairments</td>
<td>Self-report</td>
<td>Construct *Exploratory factor analysis</td>
<td>Internal consistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cronbach’s $\alpha = 0.64$</td>
</tr>
<tr>
<td>Rimmer et al., 2001</td>
<td>103 adults with disabilities and other health impairments</td>
<td>Self-report</td>
<td>Construct, concurrent, and predictive (Oxygen uptake) Multiple scores</td>
<td>Internal consistency: Cronbach’s $\alpha = 0.67 - 0.71$ Interrater: 0.92 – 0.99 Test-retest: 0.78 – 0.95</td>
</tr>
<tr>
<td>Rimmer et al., 1999</td>
<td>50 African American women with orthopedic and other health impairments</td>
<td>Self-report</td>
<td>Content</td>
<td>Internal consistency, interrater and test-retest Multiple scores</td>
</tr>
<tr>
<td>van der Ploeg et al., 2007</td>
<td>45 adults with orthopedic and other health impairments</td>
<td>Self-report</td>
<td>Criterion (Accelerometers)</td>
<td>Test-retest Spearman $r = 0.77$</td>
</tr>
<tr>
<td>Washburn et al., 2002</td>
<td>372 adults with orthopedic impairments</td>
<td>Self-report</td>
<td>Content and construct *Factor analysis</td>
<td>Internal consistency Cronbach $\alpha = 0.37 - 0.65$</td>
</tr>
</tbody>
</table>

*(continued)*
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Measure(s)</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beets, Combs et al., 2007</td>
<td>18 youth with a single or multiple disabilities</td>
<td>Pedometer</td>
<td>Criterion (Video recording) Multiple scores¹</td>
<td>Interinstrument, Interobserver agreement (IOA) Multiple scores¹</td>
</tr>
<tr>
<td>Beets, Foley, Tindall, &amp; Lieberman, 2007</td>
<td>35 youth with visual impairments</td>
<td>Talking Pedometer</td>
<td>Concurrent (NL 2000 pedometer) Multiple scores¹</td>
<td>Interunit Multiple scores¹</td>
</tr>
<tr>
<td>Motl, McAuley, Snook, &amp; Scott, 2005</td>
<td>23 adults with other health impairments</td>
<td>Pedometer</td>
<td>Criterion (Hand tally counter and video recording) Multiple scores¹</td>
<td>Not specified</td>
</tr>
<tr>
<td>Stanish, 2004</td>
<td>20 adults with mental retardation (MR)</td>
<td>Pedometer</td>
<td>Criterion (Hand-held counter) ICC = 0.95+ *ANOVA analysis</td>
<td>Interinstitute: r = 0.98 – .99 IOA: 99%</td>
</tr>
<tr>
<td>Busse, Pearson, Van Deursen, &amp; Wiles, 2004</td>
<td>Reliability- 10 healthy and 10 individuals with other health impairments; Validity- 30 healthy participants and 30 with other health impairments</td>
<td>Accelometer</td>
<td>Criterion (Movement checklist) r = 0.45, p = 0.001</td>
<td>Day to day, week to week ICC = 0.86</td>
</tr>
<tr>
<td>Bussman, Reuvenkamp, Veltink, Martens, &amp; Stam, 1998</td>
<td>4 men with orthopedic impairment and 4 age-matched men without disability</td>
<td>Accelometer</td>
<td>Concurrent: 90% agreement Predictive: 80% + (Video recording)</td>
<td>Not specified</td>
</tr>
<tr>
<td>Kozub, 2003</td>
<td>7 individuals with MR</td>
<td>Triaxial Accelerometer</td>
<td>Concurrent (Direct observation) r = 0.76</td>
<td>Not reported</td>
</tr>
<tr>
<td>Kozub, Oh, &amp; Rider, 2005</td>
<td>19 youth with visual impairments</td>
<td>Triaxial Accelerometer</td>
<td>Concurrent (Direct observation) R = 0.89, p &lt; .05</td>
<td>Interinstitute R = 0.98, p &lt; .001</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Measure(s)</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweedy &amp; Trost, 2005</td>
<td>14 adults with traumatic brain injuries (TBI)</td>
<td>Accelerometer</td>
<td>Criterion (Indirect calorimetry) $r = 0.74, p &lt; .05$</td>
<td>Not specified</td>
</tr>
<tr>
<td>van den Berg-Emons, Bussman, Balk, &amp; Stam, 2000</td>
<td>10 adults with other health impairments</td>
<td>Accelerometer</td>
<td>Concurrent 90% agreement (82–97%) Predictive: 80% + (Video tape analysis)</td>
<td>Not specified</td>
</tr>
<tr>
<td>Warms and Belza, 2004</td>
<td>Phase 1 and 2–6 adults with SCI; Phase 3–16 adults with SCI</td>
<td>Accelerometer</td>
<td>Concurrent (Self-report) $r = 0.60, p &lt; .01$</td>
<td>Interinstrument $r = 0.95 – 0.96; p = .000$</td>
</tr>
<tr>
<td>Washburn &amp; Copay, 1999</td>
<td>21 individuals with orthopedic impairments</td>
<td>Accelerometer</td>
<td>Criterion (Indirect calorimetry) Multiple scores¹</td>
<td>Interinstrument $r = 0.42$</td>
</tr>
<tr>
<td>Faison-Hodge &amp; Porretta, 2004</td>
<td>8 children with MR and 38 children without disabilities</td>
<td>Direct observation</td>
<td>Concurrent (Heart rate monitoring) PE: $r = 0.81$ (0.72–0.86) Recess: $r = .69$ (0.06–0.90)</td>
<td>IOA 94.4% (82–100%)</td>
</tr>
<tr>
<td>Taylor &amp; Yun, 2006</td>
<td>11 children with MR</td>
<td>Direct observation</td>
<td>Concurrent (Accelerometer) SOFIT: $r = 0.10$ CARS: $r = 0.61$</td>
<td>Generalizability theory (G-Theory) $\Phi = 0.98$ (SOFIT) $\Phi = 0.75$ (CARS)</td>
</tr>
<tr>
<td>Gosney, Scott, Snook, &amp; Motl, 2007</td>
<td>196 adults with MS</td>
<td>Self-report Pedometer Accelerometer</td>
<td>Unified Multiple scores¹</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Description</th>
<th>Measure(s)</th>
<th>Validity</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motl, McAuley, Snook, &amp; Scott, 2006</td>
<td>30 adults with MS</td>
<td>Self-report, Pedometer, Accelerometer</td>
<td>Unified Multiple scores(^1)</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Pan &amp; Frey, 2005</td>
<td>30 children with autism, 26 mothers, and 24 fathers</td>
<td>Accelerometer, Self-report</td>
<td>Criterion (for self-report) (r = 0.31 - 0.47)</td>
<td>Test-retest (for self-report) (r = 0.73 - 0.94)</td>
</tr>
<tr>
<td>Temple, Anderson, Walkley, 2000</td>
<td>6 adults with MR</td>
<td>Accelerometer, Self-report</td>
<td>Concurrent for self-report (Accelerometer) (ICC = 0.83)</td>
<td>Not specified</td>
</tr>
<tr>
<td>Temple &amp; Walkley, 2003</td>
<td>37 individuals MR</td>
<td>Self-report, Accelerometer</td>
<td>Concurrent for self-report (Accelerometer) (ICC = 0.78)</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

\(^1\) Reader can refer to the actual study for more information.

\* Researchers did not conduct correlation analysis and/or did not provide a single score but multiple scores.
Temple & Walkley, 2003); and (c) accelerometers and direct observations (Temple, Anderson, & Walkley, 2000). Among these, three studies reported validity evidence (criterion-related) for one of the two measures used, while two studies reported validity evidence as a unified entity was reported for pedometers, accelerometers, and self-reports. Reliability was specified in only one study (i.e., Pan & Frey, 2005). Participants in the studies using multiple measures of physical activity included (a) adults with mental retardation (two studies), (b) adults with other health impairments (two studies), and (c) children with autism spectrum disorders (one study).

**Individuals With Disabilities**

Among the reviewed studies, the most common studied classifications of individuals with disabilities included (a) orthopedic impairments (10 studies, 35.7%), (b) other health impairments (9 studies, 32.1%), and (c) mental retardation (6 studies, 21.4%). Of these, 20 studies (71.4%) targeted adults. Based on studies that met all inclusion criteria, no studies were found for individuals exhibiting deafness, deaf-blindness, learning disabilities, speech and language impairments, or serious emotional disturbances. It is important to note that among the studies reviewed focusing on individuals with mental retardation, some studies used samples including participants with Down syndrome (DS). Although studies using samples of individuals with DS were grouped together as part of studies on mental retardation (MR) and physical activity, it is important to acknowledge that individuals with DS and MR have differences that may affect physical activity and sedentary behavior in unique ways. In this review, they were grouped together because IDEA was used as the frame to identify individuals with disabilities.

**Discussion and Conclusion**

Consistent with the literature in measuring physical activity among individuals without disabilities, self-reports were found to be one of the most commonly used forms of assessing physical activity among individuals with disabilities. Among the self-report measures, three were specially designed for individuals with disabilities. The advantages and disadvantages of self-reported physical activity have been reviewed extensively elsewhere (e.g., Kriska & Caspersen, 1997); however, there are methodological and conceptual considerations researchers should take into consideration when using self-reports among individuals with disabilities.

Particular to those with mental retardation, recalling physical activity may become an issue due to cognitive limitations, thus affecting the validity of the data obtained (Rimmer, 2006). As a result, some researchers have used self-reports as proxy information completed by a direct care-provider and/or family member. Although practical and a way of minimizing issues of cognition, proxy reports may not provide a true representation of an individual’s habitual physical activity because proxy respondents may not accurately report all activity performed by the individual (Rimmer, 2006; Sallis & Saelens, 2000).

Another issue is the use of Metabolic Equivalent values (METs) to estimate activity intensity. Metabolic equivalents for various physical activities were established using healthy young to middle aged adults without disabilities (Ainsworth et al., 2000), thus limiting their utility for those with disabilities (Rikli, 2000;
Shephard, 2003). Due to the heterogeneity of individuals with disabilities, the large number of disabling conditions, and the high within and between variability among individuals with the same disability (Rikli, 1997; Rimmer, 2006), Warms (2006) argued that it may not be worthwhile to establish a compendium of MET values for such groups.

While the use of direct observations was minimal, the use of accelerometers as physical activity measures were as common as self-reports. This is consistent with the literature on physical activity measurement in which the use of more objective physical activity monitoring has been advocated (Janz, 2006). Only five studies in this review reported the use of multiple physical activity measures. Yet, the use of multiple measures has been considered the best alternative to minimize the limitations innate to each of the existing measures of physical activity (Dishman et al., 2001). Sallis and Saelens (2000) suggest the use of self-reports and motion sensors to better represent habitual physical activity. Self-report measures would provide context and type of physical activity, while motion sensors would provide duration and intensity. Dishman et al. (2001) have argued that the future of physical activity measurement may be in the use of multiple measures.

Of concern is the paucity of studies reporting validity evidence. Considering that the search produced 115 studies, only 28 (24.35%) met all inclusion criteria, which included reporting validity of the measures and/or validity of the scores obtained from the measures. Among the studies failing to meet the criterion in which a measure of physical activity was validated and validity estimates reported, authors indicated that the measure or measures used were previously validated; however, they did not validate and/or provide evidence of validity for the measure(s) used among the participants in their own study. Researchers have argued against this practice by stating that once a test or measure has been validated for one population, it cannot be generalized for use with other populations (Rikli, 1997; Yun & Ulrich, 2002). Subsequent uses of an instrument should be validated again for the new sample. Thus, validity is not test-specific.

Of the studies reporting validity, criterion-related validity evidence was found to be the most commonly reported. With the lack of true gold standards to serve as criterion measures of physical activity, however, we may be limited to evaluating agreement and disagreement between measures (Tudor-Locke & Myers, 2001). Patterson (2000) adds that even in studies where it is assumed that criterion-related evidence has been provided, researchers have merely provided convergent evidence since no existing measure of physical activity can be considered to be a true criterion measure of physical activity. This is significant since identifying a criterion measure is critical in establishing criterion-related validity evidence (Yun & Ulrich, 2002).

When establishing criterion-related evidence, the identification of an appropriate criterion is a critical first step because the evidence is obtained from the degree of empirical relationship between the physical activity measure and the criterion measure’s scores (Yun & Ulrich, 2002). In addition, obtaining a representative sample of the entire population of interest is required rather than using only a subgroup of that population (Yun & Ulrich, 2002). Unfortunately, in some of the studies reviewed, the researchers failed to identify an adequate criterion and/or the sample was not representative of the population. For example, in two studies, fitness measures were used to establish criterion-related validity evidence for two self-report measures; however, physical activity cannot be equated with physical
fitness since they are two distinct constructs (Caspersen et al., 1985). Physical activity has been defined as bodily movement produced by skeletal muscle resulting in energy expenditure, while physical fitness refers to a person’s set of attributes or traits that relate to the ability to perform physical activity (Caspersen et al., 1985). Therefore, the use of these two terms interchangeably is not appropriate because they describe different concepts. Failure to use appropriate criterion measures may lead to equivocal and/or questionable results.

Energy expenditure measures have also been used (e.g., indirect calorimetry and doubly labeled water) as criterion measures. These measures represent a product of engaging in physical activity, rather than the behavior(s) itself. Physical activity, however, is a multidimensional construct comprised of four dimensions and/or domains, which include (a) duration, (b) frequency, (c) type, and (d) intensity (LaMonte et al., 2006; Montoye, Kemper, Saris, & Washburn, 1996; Welk, 2002). In studies where indirect calorimetry or doubly labeled water were used as criterion measures, the researchers may have provided evidence of measurement validity for only one physical activity dimension. Therefore, inferences and/or decisions based on the validity evidence for that single dimension may need to be used with caution when the purpose of the study was to provide validity evidence in assessing the construct of physical activity.

Two studies used a different approach to provide validity evidence. Motl et al. (2006) and Gosney et al. (2007) used a nomological network of hypothesized direction and magnitude of relationships among multiple measures of physical activity to provide evidence of the validity of scores derived from such measures. This approach is in line with the conception of validity as a unified entity (Messick, 1995) where rather than the existence of forms of validity, there are strengths of evidence of validity (Messick, 1995; Yun & Ulrich, 2002) with construct validity at the core (Messick, 1995).

Test-retest and interinstrument were the most common forms of reliability reported in the studies reviewed; however, an important methodological concern is that some researchers failed to report time elapsed between the first test administration and the retest. This is important because in many instances, participants recalled different days on the first administration of a test than on the retest. If the instrument required participants to recall the previous seven days of activity and the retest was not given on the same day as the first administration, they would not be recalling the same time period. This has been raised as a concern for researchers in physical activity measurement (e.g., Patterson, 2000; Sallis & Saelens, 2000).

Regarding specific groups of individuals with disabilities, those with orthopedic impairments, mental retardation, and other health impairments have received greater attention in the reviewed literature compared with other disabilities such as deafness, deaf-blindness, and/or multiple disabilities. Yet it is well accepted that physical activity is important for all individuals with disabilities. On the contrary, those who have received less attention reflect our limited understanding of the physical activity needs of a wide range of disabilities and our need to expand physical activity research. In addition, most of the reviewed studies have focused on adults. This is particularly interesting when an age-related trend in physical activity has been identified among individuals with and without disabilities (Caspersen, Pereira, & Curran, 1999; Longmuir & Bar-Or, 2000) where adolescence has been identified as the key period for targeting physical inactivity (Rowland, 1999; USDHHS, 2000).
In conclusion, based on the results of this review, there is a paucity of research on the physical activity of individuals with disabilities. Using a known framework for identifying disabilities, those with mental retardation, orthopedic impairments, or other health impairments have received attention, while individuals with speech and language impairments, deaf-blindness, deafness, specific learning disabilities, or serious emotional disturbances have received little to no attention. With measurement validity as a primary need in the area of physical activity research, it is important to consider moving from the traditional view of validity as a quality of a test and/or instrument to that of the process of assuring the meaningfulness of the inferences made and decision making from measurement taken on a group-specific basis (Yun & Ulrich, 2002).

There are still many unanswered questions such as what the current levels of physical activity are among individuals with disabilities and what types and amounts of physical activity may be needed to achieve health-related outcomes (Rimmer, 2006). Through estimating measurement validity in physical activity research, we may be able to better quantify progress toward meeting national physical activity objectives and to better evaluate our efforts toward reducing physical activity disparities between those with and without disabilities (Rimmer, 2006). Laporte and colleagues (1985) commented that in order for a physical activity measure to be considered a “gold standard,” it must be valid, reliable, practical, and nonreactive. To date, no existing measure has met these four criteria. Only through continuous research and proper validity procedures can we further improve existing physical activity measures. Developing valid and reliable measurement tools for individuals with disabilities will not be a simple task. Every effort must be made, however, to establish valid and reliable physical activity measures among those with disabilities so that research is meaningful and effective (Rikli, 1997; Yun & Ulrich, 2002).

Limitations

This review was limited to the use of two search engines (PubMed and SPORT-Discus) as well as a specified time period (January 1990 to December 2007). In addition, the review was limited to papers published in English only peer-reviewed journals. As such, unpublished papers from proceedings and abstracts were excluded because of convenience.

Implications for Research and Practice

One of the major findings of this review was the paucity of published studies reporting validity evidence of physical activity measures. As a result, there is a need for more systematic physical activity reviews to address measurement issues (such as quality procedures for reporting validity and reliability evidence) among individuals with disabilities. Therefore, it is recommended that researchers in physical activity of individuals with disabilities make every effort to report validity and reliability evidence of measures used in the study (Rikli, 1997).

Future research should also focus on the physical activity of individuals with various types of disabilities for which research has been minimal. As we further our understanding about the physical activity of individuals with disabilities, the use of a combination of measurement methods (i.e., multiple measures of physi-
Physical activity measurement and disability appear to offer the most promise for understanding physical activity behavior as well as helping to enhance the quality of research designs and program interventions (Dishman et al., 2001; Fox & Riddoch, 2000; Sallis & Saelens, 2000).

References


