Computer Adaptive Test Approach to the Assessment of Children and Youth With Brachial Plexus Birth Palsy

M. J. Mulcahey, Lisa Merenda, Feng Tian, Scott Kozin, Michelle James, Gloria Gogola, Pengsheng Ni

This study examined the psychometric properties of item pools relevant to upper-extremity function and activity performance and evaluated simulated 5-, 10-, and 15-item computer adaptive tests (CATs). In a multicenter, cross-sectional study of 200 children and youth with brachial plexus birth palsy (BPBP), parents responded to upper-extremity \( (n = 52) \) and activity \( (n = 34) \) items using a 5-point response scale. We used confirmatory and exploratory factor analysis, ordinal logistic regression, item maps, and standard errors to evaluate the psychometric properties of the item banks. Validity was evaluated using analysis of variance and Pearson correlation coefficients. Results show that the two item pools have acceptable model fit, scaled well for children and youth with BPBP, and had good validity, content range, and precision. Simulated CATs performed comparably to the full item banks, suggesting that a reduced number of items provide similar information to the entire set of items.

Although the incidence of brachial plexus birth palsy (BPBP) is relatively low (Okby & Scheiner, 2012), residual motor impairment can be significant, causing secondary complications that often require occupational therapy and surgical management throughout childhood and adolescence (Hale, Bae, & Waters, 2010; Kozin, 2011). Studies on BPBP have focused primarily on outcomes associated with impairment using active movement and muscle strength as primary endpoints (Curtis, Stephens, Clarke, & Andrews, 2002; Immerman et al., 2012; Luo, Chen, Zhou, Hu, & Gu, 2011; Ozben, Atalar, Bilsel, & Demirhan, 2011; Ruchelsman et al., 2011). Largely because of the lack of precise and meaningful patient-reported outcomes (PRO) instruments, little empirical knowledge is available about how occupational therapy affects upper-extremity (UE) function and activity outcomes in children and youth with BPBP.

One of the few PRO instruments shown to be responsive to treatment of BPBP is the Pediatric Outcomes Data Collection Instrument (PODCI; Daltroy, Liang, Fossel, & Goldberg, 1998). The PODCI is a 114-item, fixed-length instrument that, although not developed using a specific occupational therapy frame of reference, evaluates outcomes relevant to occupational therapy such as UE function, physical function, activity and sports, mobility, pain, and happiness; it also has a satisfaction (with treatment) domain and normative values for comparison (Hunsaker, Cioffi, Amadio, Wright, & Caughlin, 2002). The PODCI has repeatedly been shown to be sensitive to differences among orthopedic impairments and responsive to treatment in children with chronic conditions (Amor, Spaeth, Chafe, & Gogola, 2011; Lee et al., 2010; Matsumoto, Vitale, Hyman, & Royle, 2011), including those with BPBP.
(Dedini, Bagley, Molitor, & James, 2008; Huffman, Bagley, James, Lerman, & Rab, 2005; Nath, Avila, Karicherla, & Somasundaram, 2011) and in healthy children with isolated orthopedic injuries (Kunkel, Eismann, & Cornwall, 2011).

Dedini et al. (2008); Nath et al. (2011); Huffman et al. (2005); and Bae, Waters, and Zurakowski (2008) all demonstrated the benefits of the PODCI for evaluating UE function and activities in children with BPBP but also noted potential limitations with respect to the content range of the items (Dedini et al., 2008). Likewise, Lee and colleagues (2010) demonstrated responsiveness of the PODCI to orthopedic surgery in children with cerebral palsy but also found ceiling and floor effects and possible covariance with parent satisfaction. The PODCI, like most fixed-length outcomes instruments, is vulnerable to the inherent limitations of traditional measurement; it is a relatively long questionnaire with items that may not provide useful information (e.g., items that are not applicable, items the responder has no experience with), gaps in content, and ceiling and floor effects. These limitations make it difficult to evaluate outcomes and build evidence in support of occupational therapy.

The limitations of traditional outcomes instruments, such as inadequate content range, poorly defined measurement precision, and poor match between the test content and respondent ability, have catalyzed the development of a new generation of outcomes instruments, specifically computer adaptive testing (CAT). CAT technology uses large item banks and a simple form of artificial intelligence that selects appropriate items on the basis of respondents’ pattern of responses to previous items. With CAT, clients answer items that are appropriate for their particular level of function and are never asked to respond to items that are too easy or too difficult or to items that are simply irrelevant. For example, a respondent for a child who is home schooled would never have to respond to an item about the child’s ability to sit in the lunchroom, and a respondent for a child who uses wheeled mobility would not be given an item about walking. Similarly, children younger than driving age would not be asked about their ability to drive a car, nor would an adolescent be given an item about playing with toys. In this way, only the few items that provide relevant information would be used to estimate the client’s function, making for a precise, meaningful, and low-burden instrument.

CAT would be ideal for evaluating client-reported outcomes of occupational therapy. Clients would answer 5–10 highly relevant items that would provide an estimate of their function and eliminate the need for multiple fixed-length instruments. CAT administration could occur at the point of service, making data available immediately to aid in the therapeutic discussion about goal setting, or alternatively it could be completed remotely to aid in treatment planning, longitudinal follow-up, and benchmarking within and across practice environments. Equally important, a CAT approach to assessment reflects the respect for the dignity of clients inherent in occupational therapy by avoiding administration of measures that require significant time to complete and that have items that are irrelevant or degrading (e.g., asking a person if it is difficult to walk across the street when he or she can take only several steps).

As with others who are using CAT techniques to develop precise and low-burden outcome instruments, we have described the development of CAT for use with children and youth with cerebral palsy (CP) and spinal cord injury (Bent et al., 2013; Gorton et al., 2010; Haley et al., 2009, 2010; Montpetit et al., 2011; Mulcahey, Calhoun, et al., 2012). Our studies on CATs for children and youth with CP (Gorton et al., 2010; Haley et al., 2009, 2010; Montpetit et al., 2011) showed strong test–retest reliability (intraclass correlation coefficients [ICCs] > .91); strong concurrent, convergent, and divergent validity ($r_s = .59–.91$); better precision; and broader content range when compared with traditional fixed-length instruments, including the PODCI. More recently, we have reported that the items used in the CP CATs had little to no differential item functioning in children and youth with BPBP, indicating their potential usefulness for evaluating outcomes of occupational therapy or other intervention not only for children and youth with CP but also for those with BPBP (Mulcahey, Kozin, et al., 2012).

The purpose of this study was twofold. The first purpose was to examine the psychometric properties of item pools relevant to UE function and activity performance originally designed for children and youth with CP and expanded for children and youth with BPBP. The second goal was to evaluate the item banks and simulated 5-, 10-, and 15-item CATs for known-group and concurrent validity.

**Method**

**Research Design**

This study was a cross-sectional, multicenter study. One-time data collection occurred at the point of care. Each center’s research institutional review board or ethics committee approved the study. Parents provided consent, and children and youth provided assent.
Participants

A total of 200 children and youth aged 4–21 yr with a diagnosis of BPBP formed the sample of convenience and were recruited during routine office visits from five pediatric orthopedic specialty hospitals in North America. Children and youth were excluded from the study if they had an acquired brachial plexus injury sustained after birth, if their parents were unable to read and respond to items or spoke a language other than English or Spanish, or if parents and children or youth refused to provide written informed consent and assent, respectively. Although the purposes and analytical strategies differ between this study and a previous study on differential item functioning (Mulcahey, Kozin, et al., 2012), the participants were drawn from the same sample.

Data Collection and Measures

For each participant, age, sex, race, and degree of impairment were recorded. Degree of impairment was defined using the common categories for brachial plexus injuries, including C5–6 (Erb’s palsy), C5,6,7 (extended Erb’s palsy), and C5–T1 (total plexus involvement). One research assistant (occupational or physical therapist) at each site collected data. Standardization of the data collection protocol was ensured by provision of written standardized testing procedures with accompanying case report forms for data entry and by multiday webinar training. Biweekly conference calls were also conducted to monitor the study and to address any concerns with data collection protocols.

Item Banks for UE Function and Activity Performance. The UE and activity item banks were originally developed on the basis of a thorough review of existing assessments, and new items were written relevant to UE function and activity performance of children and youth. Initial reports of item bank development and calibration with children and youth with CP are described in detail elsewhere (Gorton et al., 2010; Haley et al., 2009, 2010; Montpetit et al., 2011). All of the UE and activity items were reviewed for relevance to children and youth with BPBP, several items were modified by adding the preambles “with your impaired hand,” others that were irrelevant for children and youth with BPBP were removed, and new items specifically for children and youth with BPBP were developed. The final UE item bank included 52 items, 43 developed as part of the CP item bank and 9 newly developed. The final activity item bank included 34 items, 32 developed as part of the CP item bank and 2 newly developed. The UE item bank includes items pertaining to use of the upper extremities to perform routine, everyday tasks. The activity item bank includes items pertaining to performance of typical activities in the home, school, and community; some of the items required use of the UEIs and others did not. These item banks have been published previously (Mulcahey, Kozin, et al., 2012).

The UE and activity item banks were administered to parents using a computer tablet with custom-written software that also oriented parents to the study and provided instructions for the instrument. The custom software also included filter questions to eliminate irrelevant items (e.g., sex-related items). Response choices used a 5-point rating scale that focused on difficulty and included unable to do, can do with much difficulty, can do with some difficulty, can do with a little difficulty, and can do without any difficulty.

Box and Block Test. The Box and Block Test (BBT; Mathiowetz, Federman, & Wiemer, 1985) assesses gross manual dexterity and was used as a performance measure comparison. Multiple psychometric studies on the BBT have been conducted, including studies that established pediatric (Mathiowetz, Federman, & Weiner, 1985) and adult (Mathiowetz, Volland, Kashman, & Weber, 1985) normative values. Reliability studies on the BBT established strong test–retest reliability, with ICC values of .96–.97 (Desrosiers, Bravo, Hébert, Dutil, & Mercier, 1994; Platz et al., 2005) and Pearson product–moment correlation ($r$) values of .93–.98 (Chen, Chen, Hsueh, Huang, & Hsieh, 2009). Studies of inter- and intrarater reliability were equally strong, with ICCs of .99 (Platz et al., 2005) and $r$ values of .99–1.00 (Mathiowetz, Federman, & Weiner, 1985). Studies have also shown moderate to strong concurrent and construct validity as evidenced by $r$ values of .43–.95. In the sample described in this study, we demonstrated that the BBT was a good predictor of BPBP impairment level (Spearman rank order coefficient = .55; Mulcahey, Kozin, et al., 2012).

For BBT testing, the child was seated in a standard chair facing the box and blocks with feet on the floor. Cubes were placed in the compartment of the test box on the same side as the hand being tested (impaired side). The total number of blocks lifted, carried, and released with control in a 1-min period was counted. The BBT scores were used to evaluate convergent validity of the UE item bank and the simulated 5-, 10-, and 15-item UE CATs with actual hand function.

Pediatric Outcomes Data Collection Instrument. The psychometric characteristics of the PODCI have been examined with various patient populations; reports indicate that it can discriminate between children and youth with BPBP and normative values ($p < .01$; Bae et al.,
2008) and within known groups of children and youth with BPBP (Huffman et al., 2005). The PODCI was administered to the parents. Scores for each of the five primary dimensions within the PODCI are scaled from 0 to 100, with 100 being the highest level of UE, physical, and activity-related function, happiness, or comfort. The PODCI global score and subscale scores were used to evaluate convergent and divergent validity of the item banks and the simulated 5-, 10-, and 15-item CATs. The PODCI UE subscale was also used as a basis for evaluating the content range of the UE and activity item banks.

Analytical Procedures

Unidimensionality. The latent test structure of both item banks was tested using confirmatory factor analysis (CFA) and exploratory factor analysis (EFA). The structure of the BPBP UE and activity item banks was also compared with the structure of the CP data using multiple-group CFA based only on common items in both the BPBP and CP item banks. In CFA, model fit is assessed by multiple fit indexes, including the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), and residual correlations. CFI and TLI values >.90 indicate acceptable fit, as do RMSEA values close to or <.06. Researchers have suggested that for most fit indexes, it is difficult to establish strict cutoff criteria (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Fan & Sivo, 2007). Three primary characteristics were used to determine the extent to which a unidimensional model adequately represented the two scales in EFA: (1) the magnitude of eigenvalues of the first factor and the ratio of the magnitude of eigenvalues between the first and second factors, (2) results from the overall model fit tests, and (3) the patterns of interitem correlations among items.

Differential Item Functioning. Differential item functioning (DIF) occurs when factors other than the ability of the person influence responses. If the response is dependent on other factors, it may lead to biased estimates. DIF was examined using ordinal logistic regression (Crane et al., 2007) to evaluate whether the pattern of item responses was influenced by group membership—that is, whether participants in different BPBP categories (C5–6; C5,6,7; C5–T1) responded to the items in the same manner. The dependent variable was the responses to an item. Three independent variables include the trait level measured with the total raw score, group membership (C5–6; C5,6,7; C5–T1), and an interaction term between trait level and group membership. The analytic strategy was to successively add trait level, group membership, and interaction term into the model in three steps, and the procedure was repeated for each item. A significant group membership effect indicates the presence of uniform DIF, and a significant interaction term indicates nonuniform DIF. The following criteria were set for DIF analysis: If the likelihood difference test was statistically significant and the $R^2$ change was >.07 for one item, that item exhibited a large DIF. If the likelihood ratio test was statistically significant and the $R^2$ change was between .035 and .070, that item exhibited moderate DIF. Otherwise, values indicated negligible DIF (Jodoin & Gierl, 2001). The DIF analysis was conducted separately for each item in the UE and activity domains.

Scale Coverage. To evaluate the matching of item content with the estimated UE and activity scores of the sample, parallel item maps were produced in which item category expected values and person scores were plotted on the same metric (mean = 50, standard deviation [SD] = 10). Thus, for each item, the expected value of each rating scale category (5 per item) was used. The content range was based on estimated locations of the item–response categories that represent the lowest and highest level of ability of the sample and was examined using item maps, item difficulty, and person difficulty plotted simultaneously on the same scale.

Validity Analyses. The ability of the full item banks and each of the simulated CAT versions (5-, 10-, and 15-item versions) to discriminate among groups of children and youth on the basis of level of impairment was evaluated by comparing the mean differences in scores across the three categories of BPBP using one-way analysis of variance tests with post hoc comparisons. To assess validity, Pearson correlations were calculated between the full UE item bank and each CAT version and the Box and Block Test and between both the UE and activity item banks and each CAT version and the PODCI global score and subscale scores.

Score Precision and Standard Error. Score precision provides an indication of how well a scale estimates each ability level across a range of ability levels. We examined the precision of the UE and activity item banks and the simulated 5-, 10-, and 15-item CATs by calculating the standard errors at each ability level. For comparison, we calculated the precision of the UE subscale of the PODCI.

Results

A total of 200 children and youth with BPBP aged 4–21 yr participated in the study. The majority of participants were in elementary school (71%), were White (73%) and non-Hispanic (76%), and were girls (61%). The study sample was representative of the impairment distribution...
of the population of children and youth with BPBP (C5–6, 46%; C5,6,7, 30%; and C5–T1, 20%; Kozin, 2011).

The confirmatory factor analysis of the items in the two banks showed acceptable model fit as evidenced by CFI = .955, TLI = .952, and RMSEA = .068 for the activity item bank (N = 34 items) and CFI = .937, TLI = .934, and RMSEA = .075 for the UE item bank (N = 52 items). In the EFA of the activity items, the first factor explained approximately 55% of the total variance, and the ratio between the eigenvalues of the first and second factors was 8.31; in the EFA of the UE items, the first factor explained 58% of the total variance, and the ratio between the eigenvalues of the first and second factors was 8.98. Results of the multiple-group CFA confirmed a common scale-level structure for the UE item bank (TLI = .983, CFI = .981, RMSEA = .077) and activity item bank (TLI = .985, CFI = .983, RMSEA = .068) for BPBP and CP populations. The average interitem correlation was .6 (SD = 0.07) for UE and .6 (SD = .09) for activity.

**Differential Item Functioning**

Three items showed DIF, two UE items (“My child can use an eraser without tearing paper” and “Using only his or her hands, my child can pull up the tab on a can of soda”) and one activity item (“Cleaning the floor with a broom and dustpan includes getting the broom and dustpan out of the closet, using the broom to sweep the floor, and then using the dustpan to collect the dirt. My child cleans the floor using a broom and dustpan”). These items were retained because of the importance of their content.

**Scale Coverage**

The items in the UE item bank covered the full range of person ability, with good distribution of items throughout the entire range. We found little ceiling effect (2%) and no floor effect for the UE item bank and little ceiling effect (3%) and no floor effect for the activity item bank (Figure 1). Comparatively, the UE subscale of the PODCI had a larger ceiling effect and gaps in content along the continuum and contained items that were too easy for the sample (Figure 2). Similar findings were noted in the comparison between the activity item bank and the UE subscale of the PODCI.

**Validity Analysis**

The full UE item bank of 52 items was able to discriminate across known groups of BPBP categories, $F(2, 197) = 40.24$, $p < .0001$. Post hoc tests yielded significant results

---

**Figure 1. Item category score distribution and sample score distribution for the activity and upper-extremity item banks.** The maps show good content range of both item banks, with the easiest items at about 30 (both scales) and the hardest items at about 75 (activity) and 68 (upper extremity) on a mean of 50 and standard deviation of 10.
across all categories. The full activity item bank of 34 items discriminated across BPBP categories, \( R(2, 196) = 15.42, p < .0001 \), with post hoc tests yielding significant results between C5–6 and C5,6,7 and between C5–6 and C5–T1 but not between C5,6,7 and C5–T1.

In addition, the 5-item, \( R(2, 197) = 30.15, p < .0001 \); 10-item, \( R(2, 197) = 31.96, p < .0001 \); and 15-item, \( R(2, 197) = 33.71, p < .0001 \), UE CAT scores and the 5-item, \( R(2, 196) = 13.04, p < .0001 \); 10-item, \( R(2, 196) = 13.51, p < .0001 \); and 15-item \( R(2, 196) = 13.77, p < .0001 \), activity CAT scores were able to discriminate among BPBP categories. In CAT simulations, the only comparison that did not show a significant difference was between C5,6,7 and C5–T1 for the 5-, 10-, and 15-item activity CATs.

Correlations were significant (\( p < .0001 \)) for the full item bank and each of the three CATs with the Box and Block Test for both left-sided BPBP (\( r = .594 \), full item bank; \( r = .528 \), 5-item CAT; \( r = .537 \), 10-item CAT; \( r = .576 \), 15-item CAT) and right-sided BPBP (\( r = .623 \), full item bank; \( r = .589 \), 5-item CAT; \( r = .592 \), 10-item CAT; \( r = .598 \), 15-item CAT). As summarized in Table 1, correlations were significant (\( p \leq .01 \); \( p \leq .05 \)) among the UE and activity full item banks and 5-, 10-, and 15-item CATs with the PODCI global scale and each subscale except for Happiness. As anticipated, correlation was strongest between the UE and activity item banks and CATs and the PODCI UE and Activity subscales.

**Score Precision and Standard Error**

Standard errors (SEs) on person ability scores based on the activity item bank and the 5-, 10-, and 15-item activity CATs performed better than the UE subscale of the PODCI (Figure 3). The activity item bank demonstrated an SE of <2.2 over a wide range of person ability levels spanning nearly 2 standard deviations above and below the mean ability of the population. Each of the CATs performed similarly, with slightly reduced range but good precision across the ability range of the sample population. Similar findings were obtained with the UE item bank and simulated CATs.

**Discussion**

A uniform outcomes instrument that contains appropriate and meaningful items, that can be administered at the point of service with low burden on clients, and that has high precision would be ideal for longitudinal monitoring of outcomes and benchmarking within and across populations and service delivery models. Toward this vision, we previously described the development, calibration, and validation of large item banks and computer adaptive tests for children and youth with cerebral palsy and spinal cord
The results of the current study suggest that the UE and activity item banks also scale well with children and youth with BPBP and show minimal DIF when examined across brachial plexus impairment levels. In addition, this study demonstrated that the UE and activity item banks correlated well with legacy measures (i.e., measures that have been used repeatedly and are considered standard), discriminated among brachial plexus impairment levels, had content range that matched the sample population, and had precision that outperformed the PODCI. The simulated 5-, 10-, and 15-item CATs performed similarly to the full item banks, indicating the potential for a CAT approach to the assessment of UE function and activity performance of children and youth with BPBP.

The scale coverage of both the UE and activity item banks was excellent and showed no gaps in content, no floor effects, and minimal ceiling effects (2.0%–2.5%). In the future, we will develop and test more difficult items.

Table 1. Concurrent and Divergent Validity of the Upper Extremity and Activity Item Banks and Simulated CATs With the PODCI Global and Domain Scores

<table>
<thead>
<tr>
<th>Item Bank</th>
<th>Global</th>
<th>Upper Extremity Transfer and Mobility</th>
<th>Sports and Physical Function</th>
<th>Pain</th>
<th>Happiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents</td>
<td>192</td>
<td>196</td>
<td>195</td>
<td>196</td>
<td>196</td>
</tr>
<tr>
<td>Full item bank</td>
<td>.501 (.000)a</td>
<td>.522 (.000)a</td>
<td>.220 (.002)</td>
<td>.481 (.000)b</td>
<td>.263 (.000)b</td>
</tr>
<tr>
<td>5-item CAT</td>
<td>.460 (.000)b</td>
<td>.496 (.000)b</td>
<td>.185 (.010)</td>
<td>.438 (.001)b</td>
<td>.239 (.001)</td>
</tr>
<tr>
<td>10-item CAT</td>
<td>.477 (.000)b</td>
<td>.506 (.000)a</td>
<td>.194 (.006)</td>
<td>.463 (.001)b</td>
<td>.244 (.001)</td>
</tr>
<tr>
<td>15-item CAT</td>
<td>.493 (.000)b</td>
<td>.515 (.000)a</td>
<td>.216 (.002)</td>
<td>.471 (.000)b</td>
<td>.261 (.000)b</td>
</tr>
</tbody>
</table>

Note. Cells with no superscript letter show little or no relationship, suggestive of divergent validity. With the exception of the PODCI Happiness subscale, correlations were significant (p < .01; p < .05). CAT = computer adaptive test; PODCI = Pediatric Outcomes Data Collection Instrument.

*a* values >.50 indicate a strong relationship, suggesting that the full upper-extremity item bank and 10- and 15-item upper-extremity CATs have strong concurrent validity with the PODCI global and Upper Extremity subscale scores. *b* values between .25 and .50 indicate moderate to good relationships, suggesting moderate to good concurrent validity.

*p* < .05.

Figure 3. The precision (standard error [SE]) of the activity item bank is shown (solid black line) in comparison with the 5-, 10-, and 15-item activity CATs (AA) and the PODCI Upper Extremity (UE) subscale along the spectrum of difficulty normalized to a mean of 50 and standard deviation of 10. An SE of 3.2 is equivalent to a reliability coefficient (r) of .90, and an SE of 2.2 is equivalent to an r of .95. The 10- and 15-item CATs retains the high precision and wide content range of the full item bank but are more efficient because they administer only a few highly relevant items.
to address the small ceiling effect. The UE item bank and each of the simulated UE CATs discriminated among the three brachial plexus impairment levels, providing evidence in support of discriminant validity. The activity item bank also discriminated among all three BPBP categories. The simulated activity CATs discriminated between the C5–6 group and the two other groups (C5,6,7 and C5–T1) but not between the C5,6,7 and C5–T1 groups. The inability of the activity item bank to discriminate between children and youth with C5,6,7 and those with C5–T1 motor impairment may be indicative of the nonlinear relationship between impairment and activity performance but also may be attributable to insufficient activity items that require in-hand manipulation and fine motor prehension. In the future, we will develop additional items intended to evaluate fine prehension and in-hand manipulation in an effort to establish discriminative validity of the CATs.

Concurrent validity was established, with moderate correlation with the Box and Block Test and the PODCI. As anticipated, the correlation between the UE and activity item banks and CATs and the PODCI global, UE, and sports scores was stronger than the correlation with the PODCI mobility and pain scores. The finding that the UE and activity item banks were not strongly correlated with the PODCI pain and mobility scores provides preliminary evidence of divergent validity.

The results of the CAT simulation indicate that the 5-, 10-, and 15-item models yield accurate estimates of UE function and activity performance in children and youth with BPBP. Although we plan to examine the accuracy of the CATs in future studies by administering the full item bank along with the CAT in a prospective clinical situation, the results from the simulated CATs build evidence in support of a computer adaptive approach to evaluating children and youth with BPBP.

Additional steps are planned for the item banks and CATs described in this study. Although the initial psychometric analysis showed unidimensionality of the item banks and evidence of validity, we did not evaluate reliability. We have established good reliability of the item banks with children and youth with CP (Haley et al., 2010), but we cannot assume that the data from the item banks are reliable with children and youth with BPBP. A necessary step is a study of test–retest reliability using a prospective cohort of children and youth with BPBP. Also, a prospective, longitudinal responsiveness study is planned to evaluate the degree to which the UE and activity CATs detect change as a result of occupational therapy, orthopedic intervention, and a combination of both. In their current form, the UE and activity item banks and CATs are parent-reported outcomes of a child’s function. Ideally, we would like to examine the item banks with a child response scale as we have successfully done with CATs developed for children and youth with spinal cord injury (Mulcahey, Calhoun, Riley, & Haley, 2009). Having both a parent and a child response scale would allow evaluation of outcomes from both perspectives and would address the potential covariance described by Lee et al. (2010). Further work is warranted on how well the UE and activity item banks evaluate shoulder function and performance of activities that require overhead reach and shoulder external rotation. Finally, work is under way to establish meaningful score reports based on CAT administration to parents, children and youth, and service providers.

Limitations

This study has several limitations. First, although we enrolled sufficient participants to examine DIF between diagnostic groupings (CP vs. BPBP) and within the population with BPBP (three impairment categories), we did not have sufficient participants in each age group to examine DIF as a function of age. Additionally, the UE and activity item banks used in this study were designed as parent reports as opposed to child reports. Results with child responders may or may not be similar to the results of this study. Finally, this study describes the first of several necessary steps in the development of a patient-reported outcome. Additional psychometric work is needed to establish the test–retest reliability and to evaluate if the item banks are responsive to change in children and youth with BPBP.

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- Large item banks and CATs can potentially address the limitations of traditional instruments used in occupational therapy by minimizing or eliminating ceiling and floor effects, improving score precision, reducing the burden associated with long instruments, and generating test scores in real time.
- The results of this study demonstrated that item banks of UE function and activity performance scaled well for children and youth with BPBP and had good validity, content range, and precision.
- Simulated CATs performed comparably to the full item banks, suggesting that a reduced number of items provide similar information as the entire set of items.
Acknowledgments

Stephen M. Haley was instrumental in the development of the item banks and CATs and was coinvestigator of this study at the time of his death. We acknowledge the assistance of Leah Bent, Ross Chafetz, Judi Linton, Beth Lynch, and Kathleen Montpetit with data collection. This study was funded by a Shriners Hospitals for Children research grant (M. J. Mulcahey, principal investigator).

References


