

Clinical Study

Assessing the Discriminative Ability and Internal Consistency of the School Outcomes Measure

Sandra H. Arnold,¹ Thubi H. A. Kolobe,¹ David M. Thompson,² and Lindsay Boeckman²

¹ Department of Rehabilitation Sciences, College of Allied Health, University of Oklahoma Health Sciences Center, 1200 N Stonewall Avenue, P.O. Box 26901, Oklahoma City, OK 73126-0901, USA

² Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, 801 N.E. 13th Street, P.O. Box 26901, Oklahoma City, OK 73126-0901, USA

Correspondence should be addressed to Sandra H. Arnold; sandra-arnold@ouhsc.edu

Received 31 January 2013; Accepted 4 March 2013

Academic Editors: M. P. Galea, H. Houdijk, and H. Unalan

Copyright © 2013 Sandra H. Arnold et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The School Outcomes Measure (SOM) measures the outcomes of students who receive school-based occupational and physical therapies in the USA. This study examined the SOM's discriminative ability and internal consistency. Descriptive data from a previous study of 73 students, classified by gross motor function classification (GMFCS) level of disability, was computed to determine the frequency of use of the SOM items and differences in subscale scores by students with various ages and levels of disability. There were no differences in mean subscale scores based on age; however students with less severe disabilities (GMFCS I–III) had higher mean scores in all subscales except expresses learning all students and behavior. Cronbach's alpha coefficient was used to examine the internal consistency of items of the SOM. The correlations between many of the items within the subscales were high (.87–.99). Lower alpha coefficients were noted when the SOM was applied to students in GMFCS Levels II and III on two subscales when compared to GMFCS Levels I, IV, and V. On the basis of this evaluation, we revised the SOM to prepare it for a national field testing to measure its construct validity.

1. Introduction

Both individual and program outcomes measurement are important for determining effectiveness of student interventions and to provide occupational therapists and physical therapists, who work in the school setting, with information to make decisions about treatment approaches and program planning. Individual tools measure outcomes of individual students over time [1]. Program outcome measures compare outcomes over time in groups of students at local, state, or national levels. Used in multivariable models, program outcome measures can also identify variables with which various outcomes are associated, such as service delivery models, types of intervention, child characteristics (e.g., age, sex, and diagnosis), and service intensity [2]. Although school-based occupational therapists and physical therapists in the USA are expected to measure the effectiveness of their interventions with the students they serve [3–5], no tool currently exists that measures either individual or program

outcomes of students with disabilities who receive school-based occupational therapy or physical therapy services. Nor does a minimal data set exist for school-based therapists to identify variables of service delivery, type and intensity of intervention, and student and therapist characteristics that may be associated with student outcomes.

The School Outcomes Measure (SOM) is a program outcome measure of occupational therapy and physical therapy services provided to students with disabilities in school settings [6]. The SOM, a minimal data set, gathers information about a large number of students on an ongoing basis, using the fewest number of items. The tool takes about 10 minutes to administer when a therapist is familiar with the student and requires no manipulatives or supplies other than the test form. The SOM includes 30 functional ability items that cover seven areas (subscales) of a student's ability to fulfill roles and complete tasks in school: self-care; mobility; assuming a student's role (for preschool/elementary school students and secondary students); expressing learning

(for all students and those unable to use speech as primary means of communication); and behavior. These seven areas are addressed by school-based therapists and are based on the conceptual model described by Bundy [7] and her colleagues. McEwen et al. [6] further describe the constructs underlying these subscales.

Occupational therapists and physical therapists rate, through direct or by report, each of the 30 functional ability items on a 6-point scale according to the amount of assistance (from 1 = total assistance to 6 = independent) a student requires for each activity. The SOM also collects data on student and therapist demographics, pertinent information on the student's Individual Education Program (IEP) (e.g., disability category, educational placement, and frequency of therapy), therapy activities used (e.g., activities of daily living, assistive technology, and therapeutic interventions), and services provided, which are variables likely to be associated with student outcomes.

Research supports content validity and interrater reliability of the SOM [6]. Arnold and McEwen [8] further established item test-retest reliability of the SOM and indicated that the tool was responsive to change in students with mild/moderate functional limitations but less sensitive to change in students with severe disabilities.

However, research has not investigated if the SOM can discriminate students' performance in school based on the student's age or level of disability. Additionally, studies using the SOM have thus far used its total raw score without validating its subscales. Because the seven subscales assess diverse and sometimes unrelated tasks and activities, the use of a total raw score is questionable.

This study represents an important step in validating the SOM by assessing its discriminative ability and examining the internal consistency of its subscales. According to Portney and Watkins [9], items within each subscale should be internally consistent; that is, they should reflect "the extent to which items measure various aspects of the same characteristic and nothing else" (page 71). Evidence of internal consistency would allow the use and interpretation of the subscales scores and enhance therapists' confidence that the items within each subscale of the SOM measure the aspects of the same function (subscales).

The purpose of this study was to assess whether the mean scores on the SOM subscales differed based on student age or severity of disability and to determine the internal consistency of the functional ability items in each subscale of the SOM. In addition, we identified the frequency of use of the items and scoring criteria by students within various age groups and levels of disabilities to further refine the tool. The findings will be used in a national field test of the measure's construct validity.

2. Methods

2.1. Participants. Participants, all from the USA, were 73 students with disabilities (student participants) and 32 school-based occupational therapists and physical therapists (therapist participants) who completed the SOM for the students,

from a previous study [8]. The University of Oklahoma Health Sciences Center Institutional Review Board (IRB) approved the study. The sample of students included 49 boys and 24 girls between the ages of 3 and 20 years. Tables 1, 2, and 3 present student characteristics, primary diagnosis, and primary individuals with disabilities education act (IDEA) disability categories, respectively.

The therapists, 14 occupational therapists and 18 physical therapists, provided services through IEPs for school-aged children who received special education and related services in eight local school districts. They reported their experience in providing school-based services as follows: (a) 1-2 years of experience $n = 2$ (6.3%), (b) >2-5 years of experience $n = 5$ (15.6%), and (c) 5+ years of experience $n = 25$ (78.1%). The length of time they have worked with the student participants ranged from less than 1 year $n = 16$ (21.9%) to more than 8 years $n = 1$ (1.4%), with the majority reporting 1-3 years $n = 38$ (52.1%).

2.2. Procedures. Following a 2-hour training, provided by the first author, in classifying student disability using the Gross Motor Function Classification System (GMFCS) [10], therapist participants classified the students on their caseloads into two groups: students with mild-to-moderate functional limitations (GMFCS Levels I, II, or III) and students with severe functional limitations (GMFCS Levels IV or V). Mild/moderate and severe functional limitations were classified to initially collect response to change data [8]. Although the GMFCS criteria for classification are validated only for children with cerebral palsy, it was employed in this study because no other system exists that classifies children's gross motor function. Therapists were asked to choose, using a table of random numbers, one student from their caseload with moderate functional limitations and one student with severe functional limitations. However, one of the therapists chose only one student from the overall caseload, and three therapists selected four students for a total of 39 students with mild-to-moderate functional limitations (disability) and 34 students with severe functional limitations (disability) ($N = 73$). Therapists completed the SOM three times during one school year for each of the randomly selected student participants. This study reports item-specific response frequencies and calculates internal consistency using responses from the first administration of the SOM.

2.3. Data Analysis. We used SAS v9.2 [11] to analyze the data. We tabulated the frequency of the student and therapist demographics, pertinent student IEP information, therapy activities used, and services provided, and the mean and SD for each item's SOM scores for groups defined by age (3-5 years, 6-11 years, 12-15 years, and >16 years), and level of disability (I, II, III, IV, and V). We used the ANOVA and t -tests to determine whether the mean scores on each subscale differed based on age and severity of disability, respectively.

The Cronbach's alpha statistic was calculated to determine the internal consistency of the items within each of the SOM's seven subscales. Cronbach's alpha, the most commonly used statistical measure of internal consistency, expresses the

TABLE 1: Student participant characteristics ($N = 73$) (number (%)).

| | Total ($N = 73$) | Mild/moderate disability ($n = 39$) | Severe disability ($n = 34$) |
|---|-----------------------|--|-----------------------------------|
| Sex | | | |
| Male | 49 (67.1) | 28 (71.7) | 21 (61.7) |
| Female | 24 (32.8) | 11 (28.2) | 13 (38.2) |
| Age | | | |
| 3–5 years | 29 (39.7) | 19 (48.7) | 10 (29.4) |
| 6–11 years | 30 (41.1) | 18 (46.1) | 12 (35.2) |
| 12–18 years | 10 (13.7) | 2 (05.1) | 8 (23.5) |
| 19–20 years | 4 (05.4) | 0 (0.0) | 4 (11.7) |
| Ethnicity | | | |
| White | 54 (73.9) | 29 (74.3) | 25 (73.5) |
| African American | 15 (20.5) | 8 (20.5) | 7 (20.5) |
| Asian | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Hispanic | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Native American | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Pacific Islander | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Other: East Indian | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Gross motor function classification level (GMFCS) | | | |
| Level I | 27 (36.9) | 27 (69.2) | 0 (0.0) |
| Level II | 9 (12.3) | 9 (23.0) | 0 (0.0) |
| Level III | 3 (04.1) | 3 (07.6) | 0 (0.0) |
| Level IV | 19 (26.0) | 0 (0.0) | 19 (55.8) |
| Level V | 15 (20.5) | 0 (0.0) | 15 (44.1) |

average correlation among a set of ordinally scaled items [9]. Alpha values range from 0.00 to 1.00, with values approaching 0.90 are considered to represent high consistency [9]. We also compared values for standard Cronbach's alpha, on items within each subscale between groups defined by students' age and level of disability. Raw Cronbach's alphas were calculated when the standard deviation of any variable was 0, and there were at least four responses. We did not compute Cronbach's alpha for observations of 3 or fewer.

3. Results and Discussion

3.1. Results: Discriminative Analysis. As expected there were very little differences between the mean scores based on age (see Table 4). For all subscales, age did not have a statistically significant effect ($P > .05$). However, in all disability levels, students in disability Level I, II, and III had higher mean scores compared to students in Levels IV and V in all subscale areas except expressing learning all students ($t = -1.016$; $P = 0.3$) and behavior ($t = 1.41$; $P = 0.2$) (see Table 5). Because there were just three students classified at Level III, we further analyzed the differences between Level I and II with Levels III, IV, and V (see Table 5). Similarly, Level I and II had higher mean scores compared to students in Level III, IV, and V in all subscales except Expressing Learning All Students ($t = -0.56$; $P = 0.6$) and Behavior ($t = 1.21$; $P = 0.2$).

3.2. Results: Internal Consistency. Table 6 presents Cronbach's alpha results for all students' ratings on the SOM and shows that the values for each subscale ranged from .87 to .99. The subscales demonstrated the same high internal consistency across age and level of disabilities with a few exceptions. Tables 7 and 8 present the values for age and level of disability, respectively. The alpha values for students aged 3–5 years ranged .65–.98; .89–.99 for ages 6–11; .91–.99 for ages 12–15; and .94–1.00 for ages 15 and above. The alpha values for students across the GMFCS levels were mixed. Values for students in Level I were consistently high (.91–.99) but ranged from $-.28$ to 1.00 for the scores of students in Levels II to V. The negative correlations, although not statistically significant, were between handwriting and computer and between copying and computer ($R = -0.48$ ($P = 0.3$) and $R = -0.047$ ($P = 0.9$) resp.). As student's score increased in computer, the score decreased in handwriting or copying. The alpha values for the self-care and mobility subscales were relatively lower for those in disability Level III (.56 and .67 resp.) and the Expresses Learning All Students and Behavior subscales were low for Level II ($-.28$ and .41 resp.). Because not all therapist participants answered each item in each subscale and because several therapist participants scored the same rating on at least one of the items, both the raw and standard Cronbach's alphas were analyzed (see Table 8).

TABLE 2: Primary diagnosis of student participants ($N = 73$) (number (%)).

| | Total ($N = 73$) | Mild/moderate disability ($n = 39$) | Severe disability ($n = 34$) |
|-----------------------------------|-----------------------|--|-----------------------------------|
| Cerebral palsy | 22 (30.1) | 4 (10.2) | 18 (52.9) |
| Developmental delay | 8 (10.9) | 6 (15.3) | 2 (05.8) |
| Autism spectrum disorder | 6 (08.2) | 6 (15.3) | 0 (0.0) |
| Down syndrome | 5 (06.8) | 4 (10.2) | 1 (02.9) |
| Traumatic brain injury | 3 (04.1) | 3 (07.7) | 0 (0.0) |
| Brain tumor | 2 (02.7) | 1 (02.5) | 1 (02.9) |
| Hydrocephalus | 2 (02.7) | 0 (0.0) | 2 (05.8) |
| Anoxia | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| ADHD | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| CVA | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| DiGeorge syndrome | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Failure to thrive | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Fetal alcohol syndrome | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Lawrence-Moon | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Learning disability | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Muscular dystrophy | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Mental retardation/cerebral palsy | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Shaken baby syndrome | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Rhett syndrome | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Spina bifida | 1 (01.3) | 0 (0.0) | 1 (02.9) |
| Sturge-Weber syndrome | 1 (01.3) | 1 (02.5) | 0 (0.0) |
| Not identified | 11 (15.0) | 7 (17.9) | 4 (11.7) |

TABLE 3: Student participant primary IDEA disability category for eligibility for special education and related services ($N = 73$) (number (%)).

| | Total ($N = 73$) | Mild/moderate disability ($n = 39$) | Severe disability ($n = 34$) |
|-------------------------------|-----------------------|--|-----------------------------------|
| Multiple disabilities | 26 (35.6) | 5 (12.0) | 21 (61.7) |
| Developmental delay | 23 (31.1) | 12 (30.6) | 11 (32.5) |
| Traumatic brain injury | 5 (06.5) | 4 (10.6) | 1 (02.4) |
| Specific learning disability | 4 (05.8) | 4 (10.6) | 0 (0.0) |
| Speech or language impairment | 4 (05.8) | 4 (10.6) | 0 (0.0) |
| Autism | 3 (04.1) | 3 (07.0) | 0 (0.0) |
| Mental retardation | 3 (04.1) | 3 (07.0) | 0 (0.0) |
| Other health impairment | 3 (04.1) | 3 (07.0) | 0 (0.0) |
| Orthopedic impairment | 2 (02.4) | 1 (02.6) | 1 (02.4) |
| Deafness | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Deaf-blindness | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Emotional disturbance | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Hearing impaired | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Visually impaired (blindness) | 0 (0.0) | 0 (0.0) | 0 (0.0) |

3.3. *Results: Frequency of Responses.* The frequency data showed that all the levels of the scaling criteria for the SOM ability items were used. The exceptions were responses related to types of activities and techniques used. *Therapy Activities Used* included a rating scale of “never,” “rarely,”

“sometimes,” “frequently,” and “always” for each activity item. Analysis of these frequency counts showed that for some items, “frequently” and “always” were not selected (e.g., craniosacral therapy). Items that frequently scored “never” or “rarely” included casting/splinting, wheelchair training,

TABLE 4: Differences in subscale sum scores by age.

| Source (age group) | DF | Type III SS | Mean square | F value | Pr > F |
|---|----|-------------|-------------|---------|--------|
| Self-care | 3 | 356.8042219 | 118.9347406 | 2.72 | 0.0510 |
| Mobility | 3 | 1608.065248 | 536.021749 | 2.41 | 0.0773 |
| Assuming students role—preschool/elementary | 1 | 57.30158730 | 57.30158730 | 0.18 | 0.6799 |
| Assuming student's role—secondary | 1 | 120.3333333 | 120.3333333 | 0.27 | 0.6547 |
| Expressing learning—unable to use speech | 3 | 72.67478632 | 24.22492877 | 1.55 | 0.2198 |
| Expresses learning—all students | 3 | 137.3929737 | 45.7976579 | 2.58 | 0.0727 |
| Behavior | 3 | 245.9142424 | 81.9714141 | 2.46 | 0.0708 |

TABLE 5: Differences in sum scores by GMFCS Levels I–III versus IV–V and I–II versus III–V.

| Subscale | Estimate | Standard error | t value | Pr > t |
|--|------------|----------------|---------|---------|
| GMFCS Levels I–III versus IV–V | | | | |
| Self-care | 9.9473684 | 1.1218986 | 8.87 | <.0001 |
| Mobility | 27.447500 | 1.9138478 | 14.34 | <.0001 |
| Assuming a student's role—preschool/elementary | 25.428571 | 6.2482650 | 4.07 | 0.0016 |
| Assuming a student's role—secondary | * | * | * | * |
| Expresses learning—unable to use speech | 4.8333333 | 1.1893368 | 4.06 | 0.0003 |
| Expresses learning—all students | −1.7777777 | 1.7642292 | −1.01 | 0.3214 |
| Behavior | 2.0661696 | 1.4673178 | 1.41 | 0.1639 |
| GMFCS Levels I–II versus III–V | | | | |
| Self-care | 9.6166666 | 1.1574110 | 8.31 | <.0001 |
| Mobility | 26.757389 | 2.0214822 | 13.24 | <.0001 |
| Assuming a student's role—preschool/elementary | 25.428571 | 6.2482650 | 4.07 | 0.0016 |
| Assuming a student's role—secondary | * | * | * | * |
| Expresses learning—unable to use speech | 5.4076923 | 1.2115481 | 4.46 | <.0001 |
| Expresses learning—all students | −1.0260869 | 1.7276662 | −0.59 | 0.5569 |
| Behavior | 1.7757352 | 1.4628238 | 1.21 | 0.2292 |

*N/A as no secondary students in Levels I–III.

TABLE 6: Cronbach's alphas for all items in subscales.

| Subscale | Cronbach's alpha |
|---|------------------|
| Self-care | 0.9546 |
| Mobility | 0.9721 |
| Assuming a student's role—preschool/elementary school | 0.9897 |
| Assuming a student's role—secondary school | 0.9994 |
| Expressing learning—students unable to use speech | 0.9289 |
| Expressing learning—all students | 0.8769 |
| Behavior | 0.9380 |

craniosacral therapy, deep pressure brushing, and myofascial release. Eighty-six percent of students of age 12 years and older, were classified as having a severe disability (Level IV and V) and thus completed few subscale items related to preschool/elementary. Alternately, those classified with mild/moderate disabilities (Levels I, II, and III) completed few items in the subscale for secondary school students.

3.4. Discussion. The purpose of this study was to further refine the SOM by assessing whether the mean scores on the SOM subscales differed based on student age or severity of disability, and by determining the internal consistency of the functional ability items in each subscale of the SOM. In addition, we identified the frequency of use of the items and scoring criteria by students within various age groups and levels of disabilities.

As a minimum data set, the SOM was expected to discriminate amongst student's performance in the school setting based on level of disability. Our findings reveal differences in the mean SOM scores of the self-care, mobility, and assuming a student's role subscales, among students in Level I or II compared to Levels IV and V but not between Level I and II or Level IV and V. Although Level III was initially classified with Levels I & II, as Mild/Moderate, because there were only three students in this classification we reanalyzed the Level III data with Level IV and V data. These findings show that the SOM can discriminate among three rather than five levels of the GMFCS. One explanation may be related to the relatively few number of items in the SOM. In addition, the items of the SOM do not include the GMFCS level distinctions of need for hand-held mobility devices, wheeled mobility, or quality of movement [12]. For

TABLE 7: Standardized Cronbach's alphas for age.

| Subscale | Mean Cronbach's alpha by age group | | | |
|--|------------------------------------|----------------|-----------------|---------------------------|
| | 3–5 years old | 6–11 years old | 12–15 years old | Greater than 16 years old |
| Self-care | 0.9153 | 0.9700 | 0.9154 | 0.9459 |
| Mobility | 0.9585 | 0.9780 | 0.9533 | 0.9731 |
| Assuming a student's role—preschool/elementary school students | 0.9810 | 0.9942 | . | . |
| Assuming a student's role—secondary school students | . | . | . | 0.9994 |
| Expressing learning—students unable to use speech | 0.8018 | 0.9556 | . | . |
| Expressing learning—all students | 0.6545 | 0.8932 | . | 1.0000 |
| Behavior | 0.8985 | 0.9513 | 0.9950 | 0.9439 |

(-) signifies 3 or fewer observations; therefore, Cronbach's alpha is not computed.

TABLE 8: Standardized and (raw) Cronbach's alphas for level of disability.

| Subscale | Mean Cronbach's alpha by level of disability | | | | |
|--|--|---------|--------|----------|----------|
| | I | II | III | IV | V |
| Self-care | 0.9173 | 0.8531 | 0.5612 | 0.9190 | 0.8813 |
| Mobility | 0.9188 | 0.8093 | 0.6783 | 0.8482 | (0.6070) |
| Assuming a student's role—preschool/elementary school students | 0.9925 | . | . | . | . |
| Assuming a student's role—secondary school students | . | . | . | 1.0000 | . |
| Expressing learning—students unable to use speech | 0.9515 | 1.0000 | . | (0.0259) | (0.6116) |
| Expressing learning—all students | 0.9063 | −0.2871 | . | 0.8455 | . |
| Behavior | 0.9322 | 0.4155 | 0.9883 | 0.9547 | 0.7889 |

(-) signifies 3 or fewer observations; therefore, Cronbach's alpha is not computed.

example, distinctions between children in Levels IV and V include children having limitations in head and trunk control, requiring extensive assistive technology, and learning how to operate a powered wheelchair [12]. These specific attributes are not included within the SOM's minimal data set items. Similarly, distinctions between Levels I and II include difficulty with balance and devices/support needed for travel [12]. The SOM items do not measure quality of movement or need for devices. Thus, sensitivity or responsiveness to change may be limited if the goal of the SOM was to differentiate between the five levels of GMFCS.

On the other hand, we did not observe differences in the three subscales based on age. These findings were not unexpected given that the items in the SOM target functional rather than developmental skills. The findings support the SOM as a contextual or setting a specific tool rather than a developmental assessment scale. Similarly, the Express Learning All Students and Behavior subscales did not discriminate among students based on age or severity of disability. However, further analysis of frequency data showed that the percentages of students who exhibited independence in behavior and learning were similar in all levels except those in Level V. This group needed more support in these two areas. This suggests that the level of disability does not distinguish student's learning and behavior or the inability of the SOM to discriminate.

Outcome measures require psychometric testing to be considered reliable and valid for the population they discern. In particular, internal consistency examines if items in a tool reflect the characteristics or subscales being measured. This

research further examined the extent to which the items in each of the SOM's seven subscales were related. The results indicated that the items within the subscales of the SOM were homogeneous; that is, the items within each subscale correlated with each other, and the subscales correlated with the total scale.

Disagreement exists in the literature regarding the importance of Cronbach's alpha on establishing unidimensionality of a scale. In particular is the notion of redundancy among items leading to higher values of alpha and thus not truly guaranteeing internal consistency [13, 14]. Opponents of Cronbach's alpha as a measure of consistency base their argument on the literature from scales that measure social behavior in which evaluated behaviors could overlap. This important consideration may not apply to the SOM. First, because it is a minimal data set, efforts have been made to include a limited number of items in the SOM to capture a wide variety of areas or tasks. Second, the items are context based and assess the student's role and ability to function in the school setting and as such involves all the domains of development. As a result, a student may do well in some but not other domains regardless of age or level of disability. An example is the Behavior and mobility subscales in which the mean scores for the students in all five levels of the GMFCS were similar for the former but not for the latter.

Nunnally [15] and Nunnally and Bernstein [16] consider an alpha coefficient of .70 to represent "high consistency" in the early stages of tool development. The high alpha values observed in this study suggest that the SOM's subscales have

strong internal consistency. Although the coefficients were high, the results also suggest that some of the subscales may be more or less consistent depending on the severity of a student's disability or the assumed role expected of a student in the classroom. It should be pointed out that although Table 6 gives seven subscales, not every subscale would be appropriate for rating all students due to age and level of functioning; only five subscales would be appropriate for any given student.

Low alpha values were observed in the self-care and mobility subscales for students in the Level III category. This finding was unexpected. Previous findings from the literature on performance of children in Level III suggest that children in this category are not any more heterogeneous than the other groups. This group was underrepresented in the sample (3 students); therefore, the differences observed in our findings were more pronounced than might have been observed in a larger sample. The items in the self-care subscale such as eating, hygiene, and grooming or in the mobility subscale such as bus transfers, doors, and classroom mobility, as well as items in assuming students roles, such as obtaining classroom materials or getting teacher's attention, should not function differently for students who are in level I, II, or III [17]. The literature also suggests that where alphas are calculated for a very small number of observations (fewer than five), the alphas may not represent the subscale's performance in the larger population. Therefore, until additional studies are conducted that include a larger sample of students in Level III, these findings should be interpreted with caution.

Despite the high internal consistency of the subscales observed in the study, the inconsistency of the number of items for each subscale and the high number of subscales is a concern. For example, the mobility subscale contains eight items, and expresses learning for students unable to use speech contains only one item. An initial concern is that few items within a subscale threaten the sensitivity of scale [18]. That is, the SOM, may not be able to identify the number of students who can express learning without speech who actually can express learning, thus threatening the SOM's ability to identify positive results.

An analysis of frequency data revealed that despite the diversity of the students in terms of age and level of disability, some items were not frequently used. *Therapy activity* items such as casting/splinting, craniosacral therapy, and deep pressure brushing and subscale items such as opening and closing of lockers, going off campus for lunch, and assessing vending machines may be infrequently used in the school setting, and therefore they are not necessary items in a minimal data set. Items with frequencies of lower than 50% were eliminated from the SOM minimal data set.

In addition to removing items rarely used from the minimal data set, we expanded certain items in the functional ability subscales to distinguish specific tasks within an item. For example, in the original self-care subscale, item "eating" was expanded into two items: "eating 1" related to eating and drinking 75% of meal or snack without spilling and "eating 2" related to using a spoon and/or fork during a snack or meal. And we added two additional items in student demographics and therapy activities, respectively: state where

student resides, to collect national data, and positive behavior support, as was an "other" item frequently listed.

Finally, we changed the *therapy activities used* rating from a Likert scale to a numerical scale that accumulates percentage scores that add up to 100%. That is, each item receives a score between 0 and 100% so that the total of all items equals 100%. This change allows every item or choice to receive a value or response, including zero. All of the revisions resulted in a reduced scale that included 25 therapist, student, therapy items, and 42 student functional ability items.

The changes made to the SOM further enhance its ability to become a quick and an easy to use minimal data set for occupational therapists and physical therapists in the school setting. In addition, the continued psychometric testing establishes the SOM as a reliable and valid tool. These steps are important in developing a user friendly, valid outcome measure for practicing therapists to collect outcomes on the students they serve.

3.5. Suggestions for Further Research. Additional psychometric research is recommended to evaluate the use of the SOM as a minimal data set for measuring outcomes of students who receive occupational therapy and physical therapy services in school settings. The benefit of having a scale with good internal consistency lies in the interpretation of the score. Based on the present findings, occupational therapists and physical therapists interpret an increase in the subscale scores as an improvement in the construct measured by the subscale. However, internal consistency does not provide information about whether or not there is overlap between items in the subscales. The current seven SOM subscales were theoretically derived and determined a priori. To evaluate the validity of the subscales, it requires more advanced statistics such as factor analysis or principal component analysis [9, 19]. Additionally, Rasch model analysis can be used to create a continuous scale that will place individual SOM items along a continuum from most difficult to least difficult [20]. The importance of developing an interval scale will allow therapists to clearly report change and the magnitude of such change, when change in student's abilities actually change. Then, subsequent studies can be performed to determine the minimally clinical important difference. These analyses will require a larger sample size than that which was available in the present study.

4. Conclusion

The SOM is intended to address the need for a program measure or outcomes management system to measure the outcomes of students receiving occupational therapy and physical therapy in school settings and meet the IDEA mandate for accountability for outcomes of students with disabilities across the USA. Reliability and validity estimates are necessary to establish the SOM as a valid outcomes measure.

To assess whether the mean scores on the SOM subscales differed based on student age or severity of disability, the results of the study demonstrated that there was no difference

in mean subscale scores based on student age; however, students with less severe disabilities had higher mean scores in all subscales except expresses learning all students and behavior. In determining the internal consistency of the functional ability items in each subscale of the SOM, the study results supported internal consistency of the SOM's seven subscales, with the exception of low Cronbach's alpha for students in Levels II and III. In methodological research, correlation among all items should be determined prior to evaluation item analysis [19]. In addition, we identified frequency of use of the items and scoring criteria by students within various age groups and levels of disabilities to further refine the tool.

Therefore, the next steps in the development of the SOM are the validation studies that focus on sensitivity of each item, responsiveness of the various subscales, item difficulty [20], and the use of various scores as units of analysis. Such a process would require a large number of subjects and/or a national validation study.

Because, to date, no outcome measure exists that collects data on school-aged children, occupational therapists and physical therapists do not have a valid tool to measure change in the students they serve. A valid and reliable tool, such as the SOM, that can be used by all school-based occupational therapists and physical therapists with students of all ages and severity of disabilities will enable service providers to confidently measure the change in student ability. In addition, as a program outcome measure, the data collected from many therapists will provide national data to begin to evaluate student outcomes over time and identify variables such as type of intervention or intensity of service delivery that are associated with either positive or negative student outcomes.

Conflict of Interests

The authors declares that there is no conflict of interests.

Acknowledgment

This research was supported by funding from a grant from the Presbyterian Health Foundation.

References

- [1] E. A. Dobrzykowski, *Essential Readings in Rehabilitation Outcomes Measurement: Application, Methodology, and Technology*, Aspen, Gaithersburg, Md, USA, 1998.
- [2] W. T. F. Goossen, P. J. M. M. Epping, T. Feuth, T. W. N. Dassen, A. Hasman, and W. J. A. Van Den Heuvel, "A comparison of nursing minimal data sets," *Journal of the American Medical Informatics Association*, vol. 5, no. 2, pp. 152–163, 1998.
- [3] American Occupational Therapy Association, "Occupational therapy practice framework: domain and process," *American Journal of Occupational Therapy*, vol. 56, pp. 609–639, 2002.
- [4] American Physical Therapy Association, *Guide to Physical Therapist Practice*, American Physical Therapy Association, Alexandria, VA, USA, 2nd edition, 2001.
- [5] "Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400".
- [6] I. R. McEwen, S. H. Arnold, L. H. Hansen, and D. Johnson, "Interrater reliability and content validity of a minimal data set to measure outcomes of students receiving school-based occupational therapy and physical therapy," *Physical and Occupational Therapy in Pediatrics*, vol. 23, no. 2, pp. 77–95, 2003.
- [7] A. C. Bundy, Ed., *Making a Difference: OTs and PTs in Public Schools: A Conceptual Model of Practice for School System Therapists*, University of Illinois at Chicago, Chicago, Ill, USA, 1991.
- [8] S. H. Arnold and I. R. McEwen, "Item test-retest reliability and responsiveness of the School Outcomes Measure (SOM)," *Physical and Occupational Therapy in Pediatrics*, vol. 28, no. 1, pp. 59–77, 2008.
- [9] L. G. Portney and M. P. Watkins, *Foundations of Clinical Research: Application to Practice*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2009.
- [10] R. Palisano, P. Rosenbaum, S. Walter, D. Russell, E. Wood, and B. Galuppi, "Development and reliability of a system to classify gross motor function in children with cerebral palsy," *Developmental Medicine and Child Neurology*, vol. 39, no. 4, pp. 214–223, 1997.
- [11] S. A. S. Institute Inc, "SAS/STAT user's guide Version 8," 1999, http://books.google.com/books?id=y7gDP_ELM3gC&pg=PA1306&lpg=PA1306&dq=SAS/STAT+user%E2%80%99s+guide,+tests+and+measures+of+agreement&source=bl&ots=GI5wusNAHB&sig=LGljnLMwdFis2IRdZ5rAW7zDh98&hl=en&ei=2z13SoLfDo2QMCSbjbEM&sa=X&oi=book_result&ct=result&resnum=1#v=onepage&q=&f=false.
- [12] R. Palisano, P. Rosenbaum, D. Bartlett, and M. Livingston, *Gross Motor Function Classification System Expanded and Revised (GMFCS -E & R)*, CanChild Center for Childhood Disability Research, McMaster University, 2007.
- [13] G. J. Boyle, "Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?" *Personality and Individual Differences*, vol. 12, no. 3, pp. 291–294, 1991.
- [14] D. L. Streiner, "Starting at the beginning: an introduction to coefficient alpha and internal consistency," *Journal of Personality Assessment*, vol. 80, no. 1, pp. 99–103, 2003.
- [15] J. C. Nunnally, *Psychometric Theory*, McGraw-Hill, New York, NY, USA, 2nd edition, 1978.
- [16] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, McGraw-Hill, New York, NY, USA, 3rd edition, 1994.
- [17] D. Oeffinger, G. Gorton, A. Bagley et al., "Outcome assessments in children with cerebral palsy, Part I: descriptive characteristics of GMFCS Levels I to III," *Developmental Medicine and Child Neurology*, vol. 49, no. 3, pp. 172–180, 2007.
- [18] D. L. Streiner and G. R. Norman, *Health Measurement Scales: A Practical Guide to their Development and Use*, Oxford University Press, New York, NY, USA, 2nd edition, 2000.
- [19] M. A. Pett, N. R. Lackey, and J. J. Sullivan, *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*, Sage, Thousand Oaks, Calif, USA, 2003.
- [20] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2001.

