

Section 5: 
DEVELOPMENT **SOSI-M**

This section describes the procedures used to develop and evaluate the test items that constitute the final form of the *SOSI-M* and *COP-R*. A discussion of the item development procedures for both assessments is followed by a summary of the item analyses that were conducted for the *SOSI-M*. The final item analysis and norming sample of the *COP-R* are discussed in Section 9: *Comprehensive Observations of Proprioception-Revised*.

INITIAL DEVELOPMENT

The *SOSI-M* and *COP-R* are measures of motor performance primarily related to vestibular and proprioceptive processing. The tools were developed over the past 20 years from a foundation in the clinical observations originally described by Ayres (1972). The *SOSI-M* evolved directly from the video and instructional manual developed by Blanche (2002/2010) on the administration of structured clinical observations and the related clinical reasoning and interpretation of results.

As discussed in Section 1: Introduction, in the early 2000s, it became apparent that there was a need for a standardized measurement approach to the “clinical observations.” The *SOSI-M* was developed to provide a standardized set of administered tests with national norms that help identify vestibular and proprioceptive issues impacting postural control and motor planning. The *COP-R* was later developed to answer the need for a set of standardized observations of proprioceptive processing that could be used in conjunction with any evaluation of sensory-motor performance (Blanche, Bodison et al., 2012).

Initial Development of the *SOSI-M*

The set of tests that makes up “clinical observations” includes a wide variety of administered tasks, each of which may have multiple approaches to positioning the child, aspects of the skill being measured, and skills being measured. To develop the initial protocol, the authors based the initial 33 items and observations on an extensive review of the literature and their clinical experience. A group of four trained examiners completed independent video reviews of three cases, and two different types of analysis of interrater reliability were conducted based on the measurement utilized to rate all observations following the procedures described by Nunnally and Berstein (1994). The interrater analyses revealed that some aspects of the observations were difficult to agree on and required additional modifications and training.

SOSI-M Field Tests

Two rounds of preliminary field tests were conducted in Chile. The goals of the field tests were to test and revise the protocol, to conduct preliminary analyses of item performance, and establish reliability and validity. Additionally, the field tests provided valuable examiner feedback on administration procedures and materials. These preliminary field tests were supported by the Chilean Corporation of Sensory Integration and resulted in a preliminary Spanish Language version of the assessment.

The first field test was conducted in 2010, and included 90 typical children aged 5 through 7.11 years and 26 who were identified as having sensory integration difficulties (Blanche, Reinoso, Kiefer, & Barros, 2016). This study solidified that the administration procedures for each observation were stable across examiners, with coefficients (ICC and Kappa) ranging from 0.33 (appropriate) to 0.99 (almost perfect). Typical children also performed significantly better ($p < 0.001$) than children with sensory integration dysfunction.

The second field test was conducted in 2015 and included 212 children aged 5 through 7 years. Across both field tests, typically developing children scored significantly higher on most (but not all) items than the clinically referred sample who had been identified as presenting sensory processing or developmental challenges.

As part of the initial development process, an exploratory factor analysis was conducted on the total Chilean sample ($N = 328$) from the field tests described earlier. A three-factor solution that explained 49% of the variance was selected based on the scree plot and leveling off of the eigenvalues, along with explicit theoretical support. There were no significant difference using Oblimin and Varimax solutions. The three factors were named Postural Control, Bilateral Motor Coordination and Feedforward, and Sequencing. Table 4.1 shows the items that loaded on each of the factors.

In addition to the field tests, during this time period preliminary versions of the assessment were presented at international congresses, seminars, and continuing education courses in several countries, including Argentina, Austria, Brazil, Chile, Colombia, Ireland, Mexico, Panama, Peru, Portugal, South Africa, the United Kingdom, the United States, and Venezuela. Participating clinicians and colleagues provided valuable feedback on the utility of the assessment. At the completion of this phase, the protocol included 16 "tests" with a combination of quantitative and qualitative scoring components.

Development and Field Testing of the COP

Proprioception is multifaceted and influences behavioral regulation and motor control (Ashton-Miller, Wojtys, Huston, & Fry-Welch, 2001; Ayres, 1972, 1989; Coleman, Piek, & Livesey, 2001; Ferrell et al., 2004; Grob, Kuster, Higgins, Lloyd, & Yata, 2002; Laszlo & Sainsbury, 1993; Lephart & Fu, 2000). Children with disabilities such as autism spectrum disorder (ASD) and developmental coordination disorder (DCD) have been reported to be at higher risk for proprioceptive difficulties (Blanche, Reinoso, Chang, & Bodison, 2012; Riquelme, Hatem, & Montoya, 2016; Tseng, Tsai, Chen, & Konczak, 2018). The *COP* was originally developed as an easy-to-use observational tool for behavioral and sensory-motor functions related to proprioception that included 18 items scored using a Likert-type scale. For additional information, please refer to Blanche, Bodison et al. (2012) and Blanche, Reinoso et al. (2012) for discussion of the initial development and preliminary studies of reliability and validity.

SOSI-M AND COP PILOT VERSION

In preparation for national pilot testing in the United States, the *SOSI-M* protocol was again reviewed, taking into consideration the need for standardized materials that could be used in multiple clinical settings and the need to train examiners/data collectors to ensure standard administration and scoring. The *SOSI-M* pilot version included 34 items across 14 item sets. The *COP* protocol was also reviewed and scoring criteria for each item were updated and expanded.

The pilot versions of the *SOSI-M* and *COP* were administered to a national sample of 489 individuals aged 4-0 to 14-11, from August 2017 to January 2018. A total of 123 individuals with a range of diagnoses, including high-functioning ASD, DCD, and sensory processing disorder, were included in the pilot sample.

Item performance was tested using classical test theory (CTT) techniques. Item difficulty was assessed and items were assessed for bias using differential item functioning (DIF). No items were eliminated from the *SOSI-M* or *COP* based on these analyses; however, it was determined that there was a strong floor effect on the *SOSI-M* for 4-year-old participants. Preliminary scoring models were developed and assessed for the *SOSI-M*.

distribution of the U.S. population (U.S. Bureau of the Census, 2010). Additionally, the age distribution and basic statistics of the standardization sample are provided in Table 6.2. Please see Section 9: *Comprehensive Observations of Proprioception-Revised* for information on the *COP-R* standardization sample.

DERIVATION OF NORMS AND OTHER SCORES

As noted in Section 4: Scoring and Clinical Interpretation, raw scores, by themselves, provide little information about a child's level of performance or ability in a particular area. Meaningful interpretations of test results can be made only by using a common standard measurement system that allows individual performance to be compared with that of others of the same age. The most common and useful derived scores in educational assessment are scaled scores, standard scores, and percentile ranks. Each of these types of scores was developed for the *SOSI-M* and *COP-R* to provide a frame of reference about an examinee's performance. The methods used to develop these scores are described in this section.

Item Scores

The *SOSI-M* consists of 14 item sets and various types of tasks. Many of these tasks use different types of continuous raw data, such as time or number of completed repetitions. Other tasks rely on categorical data. To establish a standardized measurement with comparable scores, a scoring system was developed to convert raw data into a 0/1/2 scoring system. For each age, the scoring system used the median and -1 standard deviation (84% of the sample achieved the task) of the sample as the two cutoffs. In other words, if raw data are above the median, 2 points are given. If raw data are between the median and the -1 standard deviation cutoff, 1 point is given. Zero points are given if a score is below the -1 standard deviation cutoff.

The 0/1/2 scoring system was adapted and tested in both the pilot study and the norming study. The converted item scores demonstrated strong reliability (CTT analyses) and validity (receiver operating characteristics (ROC) analyses) evidence. The Cronbach alphas ranged from 0.82 to 0.91 in the pilot study and ranged from 0.82 to 0.89 in the norming study. The area under the curve (AUC) value in both pilot and norming studies was 0.81.

Appendix A, Tables A.1 and A.2, show the raw data to item score conversions for ages 5-0 through 14-11.

Scaled Scores

Scaled scores are a type of standardized score in which the distribution of raw scores has been fitted to a normal distribution with a mean of 10 and a standard deviation of 3 points. Scaled scores describe an examinee's standing relative to other individuals in the same age group in a larger normative sample and are typically used to report subtest scores. Because scaled scores provide a common

TABLE 6.1
Demographics of the SOSI-M Standardization Sample (N = 1,000)

CHARACTERISTIC	N	SAMPLE %	U.S. %
GENDER			
Female	497	49.70	48.94
Male	503	50.30	51.06
ETHNICITY			
White/Caucasian	794	79.40	75.92
Black/African American	95	9.50	14.34
Asian American	41	4.10	4.77
American Indian/Alaska Native	2	0.20	1.26
Native Hawaiian/Pacific Islander	0	0.00	0.24
Two or More Ethnicities	68	6.80	3.47
Not Reported	0	0.00	—
HISPANIC ORIGIN			
Yes	197	19.70	24.61
No	803	80.30	75.39
PARENT EDUCATION			
< High School Graduate	11	1.10	12.00
High School Graduate	131	13.10	29.96
1–3 Years College	291	29.10	27.49
Bachelor's Degree	348	34.80	19.86
Postgraduate Degree	216	21.60	10.69
Not Reported	3	0.30	—
REGION			
North Central	207	20.70	21.60
Northeast	167	16.70	17.10
South	369	36.90	37.30
West	257	25.70	24.00
METRO			
Urban/Suburban	837	83.70	81.31
Rural	163	16.30	18.69
DISABILITY			
Autism Spectrum Disorder	71	7.10	≈1.00 ^a
Developmental Coordination Disorder	26	2.60	5.00–6.00 ^a
Specific Learning Disability/Dyslexia	4	0.40	5.00–15.00 ^a
Attention-Deficit Hyperactivity Disorder	25	2.50	≈5.00 ^a
<i>Sensory Processing Disorder</i>	77	7.70	NA
<i>No Disability^b</i>	137	13.70	12.20 ^b

TABLE 6.2
Age Distribution and Basic Statistics of the
SOSI-M Standardization Sample (N = 1,000)

AGE	N	MEAN	SD
5	100	37.31	10.20
6	100	41.53	10.43
7	100	48.12	9.96
8	100	50.31	9.64
9	100	50.66	8.23
10	100	50.63	9.74
11	100	52.44	8.30
12	100	53.17	8.38
13	100	53.80	6.55
14	100	55.16	6.65

standard metric, they can also be used to compare the examinee's performance on one subtest with his or her performance on other tests that use the same scaled score metric.

Scaled scores for the *COP-R* domains were derived using methods described by Angoff (1971) in which the cumulative frequency of each raw score is computed for each age group, and the corresponding percentile ranks are plotted against the range of raw scores. The lines are then smoothed within and between age groups to remove irregularities. For each age group, new percentile ranks are read for each raw score. Corresponding z-scores are then obtained for each percentile rank, and scaled scores for the *COP-R* are calculated on the basis of a distribution having a **mean of 10 and a standard deviation of 3**.

Appendix B, Table B.1, shows the raw score to scaled score conversions for all *COP-R* domains.

Standard Scores

Standard scores are a type of standardized score in which the distribution of raw scores has been fitted to a normal distribution with a **mean of 100 and a standard deviation of 15**. Standard scores describe standing relative to others in the same age group in the normative sample and are typically used to report Overall scores. Because standard scores provide a common standard metric, they can also be used to compare the examinee's performance on one test with his or her performance on other tests that use the same standard score metric.

Standard scores were calculated for the Overall scores for the *SOSI-M* and *COP-R*. Standard score values for the *SOSI-M* were calculated from the total raw scores of all items. Standard score values for the *COP-R* were calculated from the