

RESEARCH ARTICLE



Reliability Assessment of Sensory Outcome Measure a New Developed Tool for Children with Cerebral Palsy

 OPEN ACCESS

Received: 28-09-2022

Accepted: 12-12-2022

Published: 17-01-2023

Citation: Singh S, Esht V, Agarwal A, Anandabai J (2023) Reliability Assessment of Sensory Outcome Measure a New Developed Tool for Children with Cerebral Palsy . Indian Journal of Science and Technology 16(3): 146-154. <https://doi.org/10.17485/IJST/v16i3.1948>

* Corresponding author.

shikhasinghmpt@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2023 Singh et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Shikha Singh^{1*}, Vandana Esht¹, Astha Agarwal², Jasmine Anandabai³

¹ Maharishi Markandeshwar Institute of Physiotherapy and Rehabilitation, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India

² Subharti Medical College, Swami Vivekanand Subharti University, Meerut, Uttar Pradesh, India

³ Jyoti Rao Phule Subharti college of Physiotherapy, Subharti University, Meerat, Uttar Pradesh, India

Abstract

Objective: To investigate the reliability of a recently developed Tool, designed for Hypoxic Cerebral Palsy Children (HCPC) to assess sensory deficits. **Method:** Internal consistency and all three types of reliability (intra-rater, test-retest, and inter-rater) were investigated. Twenty caregivers of HCPC were addressed. Principal rater took two readings for the intra-rater with a brief break, then another reading for test-retest after a seven-day interval. In the meantime, the observer took the readings for inter-rater reliability testing. The relative and absolute dependability of all three types of responses were evaluated after the recording. **Findings :** Spearman rank correlation and intraclass correlation values ranged from (0.934 to 1), demonstrating a very strong correlation. While the internal consistency was higher than desired, as indicated by Cronbach's alpha values, which ranged from (0.966 to 1). Cohen's kappa coefficient values for inter-rater reliability range from (0.048 to 0.188), and exhibited a small amount of agreement between the two observers. Standard Error of Measurement ranges from (1.026 to 3.810) and indicated credible results. All participant discrepancies on reliability testing were greater than the Minimal Detectable Change, indicating actual differences. **Novelty:** The uniqueness of this study lies not only in the selection of a recently developed tool for the assessment of sensory issues in HCPC, but also in its extensive analysis of all three types of reliability with both measures along with the assessment of internal consistency . It will be very helpful to researchers who want to create a new instrument or assess the precision of an existing outcome measure. It is the only study that offers a comprehensive analysis of each dependability approach together with the necessary statistical analysis. The Sensory Outcome Measure for Hypoxic Cerebral Palsy children is a reliable tool that has been developed for sensory testing in HCPC.

Keywords: CP (Cerebral Palsy); HCPC (Hypoxic Cerebral Palsy Children); SOMH (Sensory)

1 Introduction

The term "Cerebral Palsy" (CP) refers to a variety of heterogeneous, non-progressive abnormalities of the developing brain. Multidimensional evaluation is crucial to identify the underlying issues affecting children with CP. Cerebral Palsy often includes sensory problems along with motor symptoms. None of the scales exhibited good psychometric properties about sensory issues in CP till date. So reliability testing of this newly developed tool named SOMH is very important for clinical perspective⁽¹⁾. The tools for evaluating sensory difficulties in children with CP should be thoroughly standardized⁽²⁾. HCPCs exhibit many sensory problems along with increased motor issues. Sensory deficits like tactile, two-point discrimination and Proprioceptive senses were also seen in CP children and showed an adverse effect on gait. Two-point discrimination and stereotypy are both favorably correlated with motor abilities. So there is a need for a standardized and reliable tool for the assessment of sensory issues in CP children. A comprehensive assessment of sensory abnormalities in CP children will eventually result in an appropriate course of treatment, which will further enhance motor function as well⁽³⁾. Various reliability studies were also conducted earlier in the search of a reliable outcome measure for CP children but most of them solely focused on one domain like the Upper Extremity function⁽⁴⁾, assessment of balance-related tests⁽⁵⁾, motor power⁽⁶⁾, gross motor function assessment, and gait assessment⁽⁷⁾. However, other than SOMH, no scale has been developed yet for the Indian CPs to evaluate sensory impairments⁽⁸⁾.

Not only all three types of reliability (Intra rater, Inter-rater, and Test-retest) but also both measures of reliability, relative and absolute were assessed in this study which was missing in previous research⁽⁹⁾. Previous research on reliability either ignored any one of the important factors like SEM or simply examined one or two of its types^(9,10) In this study, where carryover, practice, or testing effects were not a factor, methods of intra-rater reliability can opt by the researcher. Test-retest reliability assessment methods are helpful for studies where there is a need of avoiding cumulative effects and measuring effects such as recall or training effects, as well as task familiarization errors. Inter-observer reliability is helpful in avoiding bias since it measures the extent to which various raters and observers provide consistent estimates of the same phenomenon⁽¹¹⁾. We searched for two approaches for each of the three types (i.e. relative and absolute reliability). The degree to which various measurements taken by various individuals are connected is known as relative reliability. It describes how strongly repeated measurements are related. On the other hand, the fewer people's repeated measurements depart from one another; the higher the consistency is referred to as absolute reliability. No earlier studies examined both measures for all three kinds. They were focused either on one type without proper analysis⁽¹²⁾. This study covers every factor of reliability testing that was overlooked in earlier research. All three types of reliability were taken into consideration with both measures of reliability—relative and absolute which were absent in earlier studies. This study will be highly beneficial for researchers who plan to create a new instrument or who wish to assess the accuracy of an existing outcome measure as it analyses the reliability of a newly developed tool. The aim of study was to investigate the reliability of a recently developed Tool, designed for HCPC to assess sensory deficits.

2 Methodology

This was a correlation study with a "longitudinal study design". The study was conducted in the Physiotherapy Out Patient Department (OPD) of Chhatrapati Shivaji Subharti Hospital, Swami Vivekanand Subharti University, Meerut (UP). Before the study, Ethical Approval was taken from the University Ethics Committee of Swami Vivekanand Subharti University (Ethical approval no: SMC/UECM/2021/245/153). A

total of twenty (n=20) caregivers of children with CP were recruited for the study. The technique of sampling was “Convenience sampling”. CP Children those who came to the Physiotherapy OPD were sent to the Pediatric department for screening for Hypoxic Ischemia. Both male and female children of age between 3 to 14 years were included in the study. Before the study, written informed consent was taken from the caregiver of all participants. CP children whose cause of the disease was other than hypoxic ischemia were excluded from the study. The Deaf and blind children were also not included in the sensory testing.

2.1 Procedure

The study started on September 16, 2021, and ended on October 18, 2021, with the last sample being recognized. After being screened according to the selection criteria, 20 caregivers of HCPCs were chosen for the scale's reliability testing. Written informed consent was also taken before starting the study. Anthropometric measurements such as age, height, and weight were taken for the recruited HCPC before the scale's reliability testing. Before reliability testing, the tool has already been validated. The positive and negative symptoms found in HCPC, it was separated into two parts (Part A and Part B). Both the individual portions and the total readings were examined for correlation.

(a) Intra Rater Reliability

As carryover, practice, or testing effects were not a factor, intra-rater reliability of the SOMH was established by measuring the scale in HCPC twice, with brief intervals in between to allow for fatigue from the prior session by the same rater. Sensory impairment readings were taken on the same day, with a one- or two-hour interval between them. The readings were taken on the same day by the same rater⁽¹³⁾.

(b) Test-Retest Reliability

The test-retest reliability of SOMH was calculated by the Ph.D. candidate (same rater) delivering the scale in HCPC under similar testing conditions to the previous session on two separate occasions after a seven-day break. The testing was done on the first day and then after again on the seventh day (by the same rater). This measurement theory concept called test-retest reliability measures how stable a measure is when repeated measurements are made. This is crucial because it affects how precisely we can quantify relationships with other relevant factors as well as how precisely we can characterize an object^(13,14).

(c) Inter-Rater Reliability

Inter-Rater reliability refers to the degree to which two or more persons agree. For the investigation of inter-rater reliability, two raters took readings. The Principal Investigator and one observer took readings on the first day, then after a seven-day gap, the same two raters took readings again. The Ph.D. candidate was the First rater and participated in the study as a Principal investigator. The Observer participated in the study as the second rater and took part only in the inter-rater reliability study. The second rater (observer) did not ask any questions from the caregiver and took part in the study only as an observer and put the reading only by hearing the answer of the caregiver those which were asked by the principal investigator^(15,16).

(d) Internal Consistency

When estimating a test's reliability based on a single administration, internal consistency is used. Internal consistency refers to how well a set of items assess a single construct as shown by how they vary from one another or intercorrelate. The composite score's high level of internal consistency allows the researcher to interpret it as a measure of the construct⁽¹⁷⁾.

2.2 Measures of reliability

For all three kinds, we looked for two methods:

(a) Relative reliability refers to the degree to which distinct measurements of different people are correlated. Relative reliability refers to the extent of the association between repeated measurements.

(b) Absolute reliability refers to the degree to which people's repeated measurements deviate from one another; the less they differ, the higher the reliability⁽¹⁸⁾.

3 Results and Discussion

1. Demographic Dimensions of Children recruited for Reliability Testing
2. Demographic dimensions of HCPCs according to gender difference selected for Reliability Testing. Male and female children recruited for the reliability study did not differ significantly in age (independent t-test, $p>0.05$), but were significantly different in weight and height.
3. Analysis

3.1 Intra Rater Reliability

- Relative reliability (Intra Rater)
- Absolute reliability (Intra Rater)

3.2 Test-Retest Reliability

- Relative reliability (Test Retest)

All values of the Spearman Correlation Coefficient ρ , ICC, and alpha measures were lies between (0.90 to 1), exceeding the cutoff for acceptable validity and good reliability⁽¹⁹⁾. The Cronbach's alpha value exceeds (0.8), which denotes a very high-level correlation⁽²⁰⁾.

According to Portney and Watkins, ICC values above 0.75 indicate acceptable reliability, while those below 0.75 indicate poor reliability. A score of 0.75 is considered bad to moderate. To provide adequate validity, reliability for numerous clinical metrics should exceed 0.90. All of the ICC measures were more than 0.90, indicating that they exceeded the threshold for good reliability and acceptable validity. The Cronbach's alpha value was higher than 0.8, which indicates a very good level⁽²⁰⁾.

- Absolute reliability (Test Retest)

The values of MDC vary between, 0.127 to 9.7 was indicated genuine differences⁽²¹⁾. SEM for intra-rater and test-retest reliability varies from 1.026 to 3.81.

3.3 Inter-Rater Reliability

Physical therapists frequently examine their patients in professional settings using a variety of tests and measurements. Intra-tester and inter-tester reliability is a crucial quality for therapists to feel confident using their tests and measures⁽²²⁾. Psychometric qualities should be taken into account while choosing the outcome measures⁽¹⁸⁾.

For each of the three forms of reliability testing, a total of 20 patients were recruited. As the reliability study need not involve a large number of subjects usually 15–20 subjects were enough⁽²³⁾.

We examined the two forms of reliability: relative and absolute. Previous studies merely addressed the technique of analysis or the kind of reliability, without clearly stating which method of analysis should be applied in the context of their type⁽¹⁹⁾. This study covers an overall comprehensive assessment of reliability.

It was advisable to consider both the relative and absolute reliability of a measuring scale before implementing it in clinical practice. The Spearman ρ (ρ)⁽²⁴⁾ was used to describe relative reliability, while SEM and MDC measurements were used for absolute dependability^(15,25,26).

3.4 Relative Reliability

The Spearman rank correlation coefficient (ρ) was used because the data were ordinal. In a case where, 4- or 5-point Likert-type scales are employed to measure the association between ordinal scales, the Pearson correlation is frequently used. The degree and direction of a monotonic relationship between the numbers were measured using the nonparametric Spearman correlation test, which is a variation of the Pearson correlation. By substituting the observed scores with rank scores, it is possible to get the Spearman correlation coefficient⁽²⁰⁾. The scale was split into two sections, each of the two SOMH segments (a total of 49 questions) was scored using an ordinal grading system with a range from 1 to 5 (Likert type). To express the SOMH reliability, the Spearman Rank Correlation coefficient was used⁽²¹⁾.

It is a nonparametric test. When the population is not normally distributed, when the sample size is small, and when data types like ordinal or nominal data are present, nonparametric tests perform well with skewed distributions. When analyzing these variables, nonparametric tests are the only option⁽²⁷⁾.

For all three types of reliability, the value of spearman rank correlation ranged from (0.954 to 1) which exhibits a very strong correlation^(28–30).

For evaluating the consistency of measuring scales, ICC (intraclass correlation coefficient) is advised. However, the ICC is dependent on several statistical hypotheses, such as normality and stable variance, which are rarely taken into account in applications related to health⁽³¹⁾. ICC was employed the most often for reliability assessment, especially for intra- rater and inter-observer agreements⁽³²⁾. It is also frequently used as the reliability index for test-retest, reliability^(23,31). The ICC is one of several statistical metrics that have been used to evaluate the test-retest reliability of functional connectivity⁽²⁵⁾. In general, it is utilized to analyze the consistency or conformity of two or more quantitative data⁽²⁴⁾.

To determine agreement or consistency between two evaluation methodologies, the ICC has been used. When the value is greater than 0.90, it indicates exceptional reliability and performance. Values between 0.5 and 0.75 suggest moderate reliability, 0.75 and 0.9 indicate good dependability and higher than 0.90 indicates exceptional reliability⁽³³⁾. The intra-class correlation values ranged from (0.934 to 1) indicating excellent reliability^(26,33).

Observer agreement occurs in studies that evaluate reliability. The physical therapy literature has many types of research evaluating observer agreement. The statistical approach recommended by Cohen and the corresponding reliability coefficient, Cohen kappa, is most frequently employed to evaluate observer agreement studies that involve nominal or ordinal data. The kappa paradox, which happens when observer agreement is high but the resulting kappa value is low, has recently caused Cohen kappa to come under criticism⁽²²⁾.

Inter-rater reliability was evaluated using Cohen's kappa coefficient. The importance of rater reliability comes from the fact that it shows how reliable the research's data are for the variables under consideration. It is a reliable statistic that may be applied to tests of intra-rater and inter-rater reliability. Similar to correlation coefficients, it has a range of 1 to +1, with 0 indicating complete disagreement and 1 indicating complete agreement between raters^(16,34,35). The values of inter-rater reliability with Cohen's kappa coefficient lie between (0.048 to 0.188). According to the values of the kappa coefficient, slight agreement (Kappa Paradox) was noted between the two observers⁽¹⁶⁾.

3.5 Absolute reliability

Absolute reliability is the degree by which individual frequent values deviate from one another; the lower the difference, the higher the dependability. Absolute reliability assesses the level of measurement deviation across repeated measurements. The absolute reliability was assessed using the SEM and the MDC^(36–38).

The SEM is a reliability statistic that shows how much a score varies when measured repeatedly⁽³⁹⁾. SEM for intra-rater and test-retest reliability varies from 1.026 to 3.81, indicating reliable results; as low SEM indicates more reliable results⁽⁴⁰⁾ but the values of SEM were exhibited somewhat higher for inter-rater reliability than the other two types of reliability.

The MDC is referred to as the smallest modification in a tool's values that quantifies a symptom but is not a measurement error⁽⁴¹⁾. Any change in a subject's score higher than the MDC, whether above or below the preceding score, is deemed genuine. More exactly, all of the participant discrepancies on repeated testing of test-retest and intra-rater were larger than the MDC, it was indicated for genuine differences, and for inter-rater 80% of participants exhibited actual differences⁽²¹⁾.

3.6 Internal Consistency

The estimate of internal consistency reliability in research that is unquestionably most frequently reported is Cronbach's coefficient alpha⁽²⁶⁾. Internal consistency of measure, often known as homogeneity, is a metric that can be used to assess an item's stability on a scale or in a measurement. If the scale is reliable, the score will remain consistent regardless of the order in which the items are presented. It is a commonly used metric for determining internal consistency. As a result, it was used to assess the consistency of test results across items⁽⁴²⁾. The values of Cronbach's alpha ranged from (0.966 to 1) indicating a higher than desirable value (i.e. more than 0.7) for internal consistency^(27,28,43).

3.7 Limitations

1. The number of HCPCs who took part in the SOMH reliability research was small. Low prevalence, Time of spreading of Corona Virus (Social Distancing), and caregiver negligence (for physiotherapy treatment and follow-up) were the main reasons for the limitations of this tool. Even though the sample size was smaller, it was well within the Donner et al. suggested level.
2. All three types of reliability tests were carried out at a single center. As a result, it's important to be cautious when interpreting reliability. Experts from several geographical locations, however, participated in the scale content validation.
3. All three types of reliability studies were completed on the same sample, even though readings were only taken three times, giving the caregiver a lower chance of recalling the response.

3.1.1 Strengths of the study

1. All three categories of reliability have been investigated and found to be consistent, indicating that the instrument is quite reliable.
2. Inter-rater reliability was carried out with the aid of another investigator which eliminate the possibility of assessor bias.
3. Both two methods were checked for all three types (relative and absolute reliability)

- Before interpreting the final results, internal consistency was also tested in addition to all three categories to ensure that no criteria were overlooked.

3.1.2 Further recommendations

- Ordinal scale could be converted into an interval scale so that percentage scoring could be calculated for each Domain.
- Study can be done on a large sample size
- Documentation may be done separately for each of the domains of the tool

4 Conclusion

SOMH, a recently developed tool exhibited excellent test-retest, intra-rater, and inter-rater reliability along with excellent internal consistency. This is a reliable tool that can be used in the Indian HCPC, for measuring sensory impairments.

Table 1. Description of Three types of Reliability

Intra Rater	Test Retest	Inter Rater
<ul style="list-style-type: none"> Assessment Session 1 (Principal Investigator) Short Period Break Assessment Session 2 (Principal Investigator) 	<ul style="list-style-type: none"> Assessment Session 1 (Principal Investigator) 7 Days Break Assessment Session 2(Principal Investigator) 	<ul style="list-style-type: none"> Assessment Session 1(Principal Investigator) Observational evaluation 1(Observer) 7 Days Break Assessment Session 2 (Principal Investigator) Observational evaluation 2 (Observer)

Table 2. Demographic dimensions of the HCPs recruited for Reliability testing (n=20)

Demographic dimensions	Mean	Range
Age (years)	5.62	3 to 14
Height (cm)	104.30	89 to 160
Weight (kg)	14.97	12 to 48

Table 3. Male and Female HCP, s selected for Reliability study demographic data

Demographic dimensions	Male (n=12)	Female (n=8)	p-value*
Age (years)	8.83	5.62	0.066
Height (cm)	125.35	104.03	0.049
Weight (kg)	27.83	14.97	0.026

Demographic dimensions of male and female HCP recruited for Reliability testing (n=20)

Table 4. Spearman Correlation Coefficient rho (ρ), Intra class coefficient (ICC) correlation coefficient and Cronbach's alpha in reporting intra-rater reliability and internal consistency of SOMH among the patient with HCP

Domains	Session – 1	Session – 2	Spearman rho (ρ)	ICC (3,k) 95% CI	Cronbach's α
PART A	87.20(30-150)	84.10(29-145)	0.998	0.999	0.999
PART B	52.15(19-95)	51.50(20-95)	0.964	0.992	0.996
TOTAL	139.35(49-245)	138.65(54-245)	0.995	0.999	0.998

All values of the Spearman Correlation Coefficient rho (ρ), ICC and alpha measures were greater than (0.90), exceeding the cutoff for acceptable validity and good reliability⁽¹⁵⁾. The Cronbach's alpha value exceeds (0.8), it denotes a very high level correlation⁽¹⁶⁾.

Table 5. SEM and MDC for two sessions of intra rater reliability

Domains	Session – 1	Session – 2	SEM	MDC
PART A	87.20(30-150)	84.10(29-145)	3.34	9.25

Continued on next page

Table 5 continued

PART B	52.15(19-95)	51.50(20-95)	1.026	2.84
TOTAL	139.35(49-245)	138.65(54-245)	3.81	10.56

The values of SEM lies between (1.026to 3.81)indicated reliable result ,as low SEM indicates more of the reliable result⁽¹⁷⁾.

Table 6. Spearman Correlation Coefficient rho (ρ), Intra class coefficient (ICC) correlation coefficient and Cronbach's alpha in reporting Test Retest reliability and internal consistency of SOMH among the patient with HCP

Domains	Session – 1	Session – 2	Spearman rho (ρ)	ICC (3,k) 95% CI	Cronbach's α
PART A	87.2(30-150)	87.0(30-150)	1.00	1	1
PART B	52.15(19-95)	52.15(20-95)	0.900	0.987	0.993
TOTAL	139.35 (49-245)	139.15 (53-245)	0.993	0.997	0.999

Table 7. SEM and MDC for two sessions of test retest reliability

Domains	Session – 1	Session – 2	SEM	MDC
PART A	87.2(30-150)	87.0(30-150)	0.046	.127
PART B	52.15(19-95)	52.15(20-95)	1.46	4.04
TOTAL	139.35(49-245)	139.15(53-245)	3.44	9.7

The values of SEM were between (1.04to 3.44) indicated reliable result, as low SEM indicates more of the reliable result⁽¹⁴⁾. The values of MD varies between (2.84 to 11) , indicated genuine differences between the two session⁽⁷⁾.

Table 8. Results of inter rater reliability between two raters for two repeated sessions

S. No.	RATER 1	RATER 2 (observer)	ICC 95% CI	Cronbach's α	MDC	SEM	KAPPA
1.	139.35 (49-245)	135.15 (59-241)	0.934	0.966	68.27	24.73	0.188
2.	139.25 (54-245)	136.45 (59-241)	0.940	0.969	65.97	23.8	0.048

All of the ICC measures were more than 0.90, indicating that they exceeded the threshold for good reliability and acceptable validity⁽¹¹⁾. The Cronbach's alpha value is higher than 0.8 , it indicates a very good level⁽¹²⁾. The values of SEM were between (23 to24.73)indicated reliable result,as low SEM indicates more of the reliable result⁽¹³⁾. The values of MD varies between (65.97to68.27) it was indicated for genuine differences⁽³⁾. The values of Cohen's kappa coefficient for inter rater reliability lies in between (0.048 to 0.188) according to values of kappa coefficient slight agreement was noted between two observers⁽⁴⁾.

References

- Singh S, Esht V, Agarwal A, Development. Validation and reliability testing of a sensory outcome measure for hypoxia induced cerebral palsy children: A study Protocol. *Turkish Online Journal of Qualitative Inquiry*. 2021;12(10):4179–4188. Available from: <https://www.tobji.net/index.php/journal/article/view/8425/5990>.
- Klingels K, De Cock P, Molenaers G, Desloovere K, Huenaearts C, Jaspers E, et al. Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably? *Disability and Rehabilitation*. 2010;32(5):409–416. Available from: <https://doi.org/10.3109/09638280903171469>.
- Akkaya KU, Elbasan B. An investigation of the effect of the lower extremity sensation on gait in children with cerebral palsy. *Gait Posture*. 2021;85:25–30. Available from: <https://doi.org/10.1016/j.gaitpost.2020.12.026>.
- Mendoza-Sánchez S, Molina-Rueda F, Florencio LL, Carratalá-Tejada M, Cuesta-Gómez A. Reliability and agreement of the Nine Hole Peg Test in patients with unilateral spastic cerebral palsy. *European Journal of Pediatrics*. 2022;181(6):2283–2290. Available from: <https://doi.org/10.1007/s00431-022-04423-w>.
- Shah P, Chauhan, C. Reliability of Early Clinical Assessment of Balance Scale in Assessment of Children with Cerebral Palsy. *International Journal of Science and Research (IJSR)*. 2018;9(8):517–520. Available from: <https://www.ijsr.net/archive/v9i8/SR20809103014.pdf>.
- Eken MM, Dallmeijer AJ, Buizer AI, Hogervorst S, Van Hutten K, Piening M, et al. Intraobserver Reliability and Construct Validity of the Squat Test in Children With Cerebral Palsy. *Pediatric Physical Therapy*. 2020;32(4):399–403. Available from: <https://doi.org/10.1097/PEP.0000000000000736>.
- Zanudin A, Khong YY, Chong LF, Mohamad NA. Test-Retest Reliability and Construct Validity of Two-Minute Walk Test in Children and Adolescents with Cerebral Palsy. *Walailak Journal of Science and Technology (WJST)*. 2021;18(11). Available from: <https://doi.org/10.48048/wjst.2021.9588>.
- Patel DR, Neelakantan M, Pandher K, Merrick J. Cerebral palsy in children: a clinical overview. *Translational Pediatrics*. 2020;9(S1):S125–S135. Available from: <https://doi.org/10.21037/tp.2020.01.01>.
- Araneda R, Ebner-karestinos D, Paradis J, Saussez G, Friel KM, Gordon AM, et al. Reliability and responsiveness of the Jebsen-Taylor Test of Hand Function and the Box and Block Test for children with cerebral palsy. *Developmental Medicine & Child Neurology*. 2019;61(10):1182–1188. Available from: <https://doi.org/10.1111/dmcn.14184>.

- 10) Dehghan L, Abdolvahab M, Bagheri H, Dalvand H, Zade F, S. Inter rater reliability of Persian version of Gross Motor Function Classification System Expanded and Revised in patients with cerebral palsy. *Daneshvar Medicine*. 2020;18(6):37–44. Available from: https://Daneshvarmed.shahed.ac.ir/article_1446.html?lang=en.
- 11) Revelle W, Condon DM. Reliability from α to ω : A tutorial. *Psychological Assessment*. 2019;31(12):1395–1411. Available from: <https://doi.org/10.1037/pas0000754>.
- 12) Daniel LF, Reina R, Gorla JI, Bastos T, Roldan A. Validity and Reliability of a Test Battery to Assess Change of Directions with Ball Dribbling in Parafootballers with Cerebral Palsy. *Brain Sciences*. 2020;10(2):74. Available from: <https://doi.org/10.3390/brainsci10020074>.
- 13) Schwartz AH, Albin TJ, Gerberich SG. Intra-rater and inter-rater reliability of the rapid entire body assessment (REBA) tool. *International Journal of Industrial Ergonomics*. 2019;71:111–116. Available from: <https://doi.org/10.1016/j.ergon.2019.02.010>.
- 14) Noble S, Scheinost D, Constable RT. A guide to the measurement and interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences*. 2021;40:27–32. Available from: <https://doi.org/10.1016/j.cobeha.2020.12.012>.
- 15) Grgic J, Lazinica B, Schoenfeld BJ, Pedisic Z. Test–Retest Reliability of the One-Repetition Maximum (1RM) Strength Assessment: A Systematic Review. *Sports Medicine - Open*. 2020;6(1):1–16. Available from: <https://doi.org/10.1186/s40798-020-00260-z>.
- 16) Mchugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22(3):276–282. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052>.
- 17) Peyton C, Pascal A, Boswell L, Deregnier R, Fjortoft T, Støen R, et al. Inter-observer reliability using the General Movement Assessment is influenced by rater experience. *Early Human Development*. 2021;161:105436. Available from: <https://doi.org/10.1016/j.earlhumdev.2021.105436>.
- 18) Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy*. 2000;86(2):61211–61215. Available from: [https://doi.org/10.1016/S0031-9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4).
- 19) Lewis JS, Valentine RE. The pectoralis minor length test: a study of the intra-rater reliability and diagnostic accuracy in subjects with and without shoulder symptoms. *BMC Musculoskeletal Disorders*. 2007;8(1):1–10. Available from: <https://doi.org/10.1186/1471-2474-8-64>.
- 20) Ursachi G, Horodnic IA, Zait A. How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. *Procedia Economics and Finance*. 2015;20:123–132. Available from: [https://doi.org/10.1016/S2212-5671\(15\)00123-9](https://doi.org/10.1016/S2212-5671(15)00123-9).
- 21) Weir JP. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *The Journal of Strength and Conditioning Research*. 2005;19(1):231–231. Available from: <https://doi.org/10.1519/15184.1>.
- 22) Park H. Reliability using Cronbach alpha in sample survey. *The Korean Journal of Applied Statistics*. 2021;34(1):1–8. Available from: <http://dx.doi.org/10.5351/KJAS.2021.34.1.001>.
- 23) Donner A, Eliasziw M. Sample size requirements for reliability studies. *Statistics in Medicine*. 1987;6(4):441–448. Available from: <https://doi.org/10.1002/sim.4780060404>.
- 24) Schober P, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*. 2018;126(5):1763–1771. Available from: <https://doi.org/10.1213/ANE.0000000000002864>.
- 25) Chamorro C, Arancibia M, Trigo B, Arias-Poblete L, Jerez-Mayorga D. Absolute Reliability and Concurrent Validity of Hand-Held Dynamometry in Shoulder Rotator Strength Assessment: Systematic Review and Meta-Analysis. *International Journal of Environmental Research and Public Health*. 2021;18(17):9293. Available from: <https://doi.org/10.3390/ijerph18179293>.
- 26) Castro-Luna G, Jiménez-Rodríguez D. Relative and Absolute Reliability of a Motor Assessment System Using KINECT® Camera. *International Journal of Environmental Research and Public Health*. 2020;17(16):5807. Available from: <https://doi.org/10.3390/ijerph17165807>.
- 27) Doorn JV, Ly A, Marsman M, Wagenmakers EJ. Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ . *Journal of Applied Statistics*. 2020;47(16):2984–3006. Available from: <https://doi.org/10.1080/02664763.2019.1709053>.
- 28) Looney MA. When Is the Intraclass Correlation Coefficient Misleading?. 2009. Available from: https://doi.org/10.1207/S15327841Mpee0402_3.
- 29) Akoglu H. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*. 2018;18(3):91–93. Available from: <https://doi.org/10.1016/j.tjem.2018.08.001>.
- 30) Song HY, Park S. An Analysis of Correlation between Personality and Visiting Place using Spearman's Rank Correlation Coefficient. *KSII Transactions on Internet and Information Systems*. 2020;14(5):1951–1966. Available from: <https://doi.org/10.3837/tiis.2020.05.005>.
- 31) Cibulka MT, Strube MJ. The Conundrum of Kappa and Why Some Musculoskeletal Tests Appear Unreliable Despite High Agreement: A Comparison of Cohen Kappa and Gwet AC to Assess Observer Agreement When Using Nominal and Ordinal Data. *Physical Therapy*. 2021;101(9):1–5. Available from: <https://doi.org/10.1093/ptj/pzab150>.
- 32) Consiglio C, Mazzetti G, Schaufeli WB. Psychometric Properties of the Italian Version of the Burnout Assessment Tool (BAT). *International Journal of Environmental Research and Public Health*. 2021;18(18):9469. Available from: <https://doi.org/10.3390/ijerph18189469>.
- 33) Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*. 2019;203:116157. Available from: <https://doi.org/10.1016/j.neuroimage.2019.116157>.
- 34) Gebrehiwet T, Luo H. Analysis of Delay Impact on Construction Project Based on RII and Correlation Coefficient: Empirical Study. *Procedia Engineering*. 2017;196:366–374. Available from: <https://doi.org/10.1016/j.proeng.2017.07.212>.
- 35) Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*. 1973;33(3):613–619. Available from: <https://doi.org/10.1177/001316447303300309>.
- 36) Rodrigues IB, Adachi JD, Beattie KA, Macdermid JC. Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders*. 2017;18(1):540. Available from: <https://doi.org/10.1186/s12891-017-1914-5>.
- 37) Johansen KL, Stistrup RD, Schjøtt CS, Madsen J, Vinther A. Absolute and Relative Reliability of the Timed 'Up & Go' Test and '30second Chair-Stand' Test in Hospitalized Patients with Stroke. *PLoS One*. 2016;11(10):165663. Available from: <https://doi.org/10.1371/journal.pone.0165663>.
- 38) Burke A, Dillon S, O'connor S, Whyte EF, Gore SF, Moran KA. Relative and absolute reliability of shank and sacral running impact accelerations over a short- and long-term time frame. *Sports Biomechanics*. 2022;14:1–16. Available from: <https://doi.org/10.1080/14763141.2022.2086169>.
- 39) Overend T, Anderson C, Sawant A, Perryman B, Locking-Cusulito H. Relative and Absolute Reliability of Physical Function Measures in People with End-Stage Renal Disease. *Physiotherapy Canada*. 2010;62(2):122–128. Available from: <https://doi.org/10.3138/physio.62.2.122>.
- 40) Schlager A, Ahlqvist K, Rasmussen-Barr E, Bjelland EK, Pingel R, Olsson C, et al. Inter- and intra-rater reliability for measurement of range of motion in joints included in three hypermobility assessment methods. *BMC Musculoskeletal Disorders*. 2018;19(1):1–10. Available from: <https://doi.org/10.1186/s12891-018-2290-5>.
- 41) Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomás M, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC Musculoskeletal Disorders*. 2008;9(1):43. Available from: <https://doi.org/10.1186/1471-2474-9-43>.

- 42) Raadt AD, Warrens MJ, Bosker RJ, Kiers H. A Comparison of Reliability Coefficients for Ordinal Rating Scales. *Journal of Classification*. 2021;38(3):519–562. Available from: <https://doi.org/10.1007/s00357-021-09386-5>.
- 43) Deutsch A, Palmer L, Vaughan M, McMullen T, Karmarkar A, Kwon S, et al. Inpatient Rehabilitation Facility Change in Self-Care and Change in Mobility Quality Measures: Development and Reliability and Validity Testing. *Archives of Physical Medicine and Rehabilitation*. 2022;103(6):1105–1112. Available from: <https://doi.org/10.1016/j.apmr.2021.12.031>.