

Supplementary table 5: Definition of terms

| | Measurement Property | Definition | Example/explanation |
|-----------------------|-----------------------------|--|---|
| Validity | Content | The degree to which an assessment tool's content measures the construct that it intends to measure ⁷ | Concerned with the relevance and comprehensiveness of the items included in the assessment tool |
| | Construct | Measures the degree to which the scores obtained from the test are an adequate reflection of the construct to be measured ⁷ | Examples include structural validity (whether scores reflect the dimensionality of the construct), hypothesis testing (item construct validity) and cross-cultural validity (whether translated or culturally adapted assessments adequately reflect the original version) ⁷ |
| | Criterion | Assesses whether or not the test scores reflect a 'gold standard' assessment ⁷ | As there is no gold standard of assessment for gross motor function in children this is often assessed with correlations of scores obtained from two or three other frequently used tools. |
| Reliability | Reliability | Refers to the consistency of a test score regardless of the time between assessments (test-retest) or the person administering (intra and inter-rater) ³⁰ | Usually measured with intraclass correlation coefficient (ICC), but can be measured using Cohen's kappa coefficient. Percentage agreement and Pearson's correlation coefficient do not incorporate error into the calculations and as such is not a true measure of agreement ⁵⁰ . Scores > 0.80 are considered excellent, 0.60-0.79 adequate and <0.59 poor ¹¹ |
| | Internal consistency | The degree of interrelatedness of an assessment tool's items ⁷ | Usually measured using Cronbach's alpha (α) ⁷ . scores > 0.70 demonstrates high relationship, 0.5 to 0.69 a moderate relationship, 0.26 to 0.49 a low relationship and < 0.26 little relationship ⁵⁰ . |
| | Measurement Error | Refers to the error obtained between measurements that cannot be attributed to the patients true change ⁷ | May be systematic or random error ⁷ |
| Responsiveness | Responsiveness | An assessment tool's ability to detect change over time in the construct it purports to measure ⁷ | This is central to a tools capacity to be used as an outcome measure. |