A synthesis of potential impairment assessment tools for Para dressage classification

Rachel C Stockley¹, Lindsay St George², Jill Alexander², Joseph Spencer¹, and Sarah Jane Hobbs²

Received: 31st May 2021; Accepted: 19th July 2022; Published: 18th November 2022

Abstract: The key aim of classification in Para sport is to try and ensure that competitors are grouped so that they compete against others with impairments that cause a similar level of activity limitation within a specific sport. This synthesis aimed to identify existing, valid, and reliable, impairment assessment tools to measure eligible impairments that influence an individual’s ability to compete at Para dressage. A multi-stage approach was employed, where a systematic search of professional databases of impairment assessment tools first identified 35 tools for Para dressage. Then, a search strategy was developed, based on these 35 tools, and 305 scientific articles were identified from academic databases up to September 2021. From here, impairment assessment tools were evaluated and refined in a two-stage process using known performance determinants for Para dressage and then an assessment of their reliability, validity and practical usability. This resulted in the selection of impairment assessment tools, which included measures of muscle tone, strength, coordination, sitting balance, and trunk function. From this synthesis, we present a novel process by which impairment assessment tools were selected, refined, and critically examined using knowledge of performance determinants for dressage, the views and experiences of stakeholders, and reliability and validity of tools. The processes described here could be applied to the development of evidence-based classification systems across a range of Para sports.

Keywords: sport; psychometric; clinical assessment; horse; rider; equestrian; athlete

Introduction

Accurate classification of Para athletes and subsequent allocation of sport class is vital to provide parity in Para sports. It ensures that Para athletes who succeed do so because of an optimal combination of training, physiological, and psychological factors, and not because they are less impaired (Connick et al., 2018a; de Jong et al., 2010). The results of these measures determine the predicted impact of impairment on sports performance and the subsequent sport class to which a Para athlete should be allocated, without which the differentiation between athletes in competition would not be achieved (Mann et al., 2021).

In 2015, the International Paralympic Committee (IPC) published the Athlete Classification Code, which advocated the development of evidence-based classification systems across Para sports (International Paralympic Committee, 2015). The IPC, and others (Connick et al., 2018a; Tweedy et al., 2014), highlighted that the classification process should use robust, transparent methods of assessment and be reliable between classifiers.
The chosen assessment tools should be refractory to training, so that Para athletes can gain a competitive advantage within their current classification by improving their performance (for example, through enhanced training and skill acquisition) but not then be re-classified based on their success. Conversely, changes in the severity of a health condition, which impact an athlete’s impairment(s) and activity limitation(s) should be detected through appropriate testing and, where appropriate, result in re-classification. The association between impairment and performance within a Para sport should be based on empirical evidence applicable to the demands of the individual sport, with each sport developing its own classification system (Mann et al., 2021).

Para dressage is the only equestrian sport that is featured in the Paralympics, and it has been included since 1996. In Para dressage, riders complete a range of movements and paces on their horse (dependent upon their grade), which are scored by judges. The competitor with the highest overall score wins. Para dressage is governed by the Fédération Équestre Internationale (FEI), has nine eligible impairments (as it does not include intellectual impairment) and classifies athletes into five sport classes, referred to as grades by the FEI (Fédération Equestre Internationale, 2022).

The current Para dressage classification system assesses eligible impairments, which are scored using ordinal level scales. Scores are then used to develop a Para athlete profile that is used to determine which of the five grades, if any, an individual should be assigned to (Fédération Equestre Internationale, 2022). A ridden observation is conducted by the classification panel during training and competition to confirm that the impairment recorded during the athlete assessment is congruent with mounted performance (Fédération Equestre Internationale, 2022).

To meet the IPC Classification Code criteria, classification methods should be objective, precise, reliable, only influenced by impairments in body structure/function, refractory to training, and ideally ratio scaled (Beckman et al., 2017; Tweedy & Vanlandewijck, 2011; Tweedy et al., 2014). To ensure that the current Para dressage classification system is based on strong scientific evidence (Fédération Equestre Internationale, 2017), objective and robust tools must therefore be identified to measure the eligible impairments that influence Para dressage performance (Tweedy et al., 2014; Mann et al., 2021).

This study aims to identify existing impairment assessment tools that could measure the impairments, which are recognized to have the greatest impact on an individual’s performance in Para dressage. These impairment assessment tools must meet specific criteria, namely, to exhibit robust properties (particularly reliability and validity) and be suitable for practical classification testing purposes.

**Materials and Methods**

A broad strategy to identify suitable tools and appraise their suitability for Para dressage was developed by three authors (RS, LSG, and SJH). This had four stages as shown in Figure 1 and described in detail in each subsection below.

---

**Figure 1.** Stages of the strategy.
Identification of potentially suitable impairment assessment tools.

All available impairment assessment tools were identified by screening two professional databases for rehabilitation tests and measures; Shirley Ryan Ability Lab (rehabmeasures.org) and PT Now (ptnow.org). Impairment assessment tools were selected if they assessed impairment in accordance with the WHO International Classification of Functioning, Disability and Health definition (World Health Organisation, 2001) and was used to assess one of the nine eligible impairments for Para dressage, as defined by the FEI (Fédération Equestre Internationale, 2022). In accordance with the current classification criteria and the positional requirements for Para dressage (Fédération Equestre Internationale, 2022), tools were only included if most items were measured in seated, supine or prone positions. Potential tools were excluded if they were patient reported outcome measures.

Impairment assessment tools were included if they produced objective, interval or ratio scaled data, as recommended by the IPC. If none of the tools for assessing an eligible impairment produced interval or ratio scaled data, then ordinal (ranking) tools were included.

Literature search to establish the psychometric properties of suitable impairment assessment tools.

Keywords were obtained by screening the peer reviewed, published articles that were included in the two professional databases (outlined above) for each assessment tool (Appendix A). If we did not find keywords within a peer reviewed, published article, we extracted the MeSH terms from Pubmed instead. These keywords and MeSH terms were grouped using the Boolean operator “OR” into four overarching themes: impairment, measures, assessment, and general terms, which were then combined using the Boolean operator “AND” to perform a systematic search (Appendix A) of online academic databases (SportDiscuss, CINAHL, MEDLINE, EMBASE and AMED) up to September 2021.

Four independent researchers (LSG, JS, RS, EHL) conducted initial title and abstract screening in EndNote. Each article was screened independently by one of the four researchers. Articles were included if they were written in the English language and assessed validity, reliability or responsiveness of an eligible impairment assessment tool within adult clinical populations (18 years or older). Grey literature, conference abstracts, and systematic reviews were excluded. When a decision to include or exclude an article could not be made between the researchers, the lead independent reviewer (Physiotherapist RS) made a definitive decision.

Further refinement of impairment assessment tools based on known performance determinants for Para dressage.

After completion of title and abstract screening, we refined the number of impairment assessment tools that would progress to the data extraction stage. We developed the refinement criteria using findings from a scoping review (Hobbs et al., 2020) and an interview study (St. George et al., 2021). These studies were conducted as part of our wider research project to define determinants of performance for Para dressage. The scoping review included 58 studies and found that the most influential characteristics of riding performance were trunk control, pelvic control, and coordination (Hobbs et al., 2020). In the interview study, 30 Para dressage stakeholders described the importance of the athlete’s ability to maintain dynamic postural control for absorbing the horse’s movement and coordinating leg, hand, and seat aids (St George et al., 2021). Taken together, rider balance, core-stability, muscle power, adequate joint range of motion, coordination, and symmetry were derived as key performance determinants for Para dressage (Hobbs et al., 2020; St. eujapa.upol.cz
George et al., 2021). Thus, we obtained full-text articles for data extraction, which assessed the reliability and validity of impairment assessment tools that measured aspects of a key performance determinant for Para dressage.

Assessment of reliability, validity, and usability of impairment assessment tools.

Six independent researchers (LSG, RS, JS, PS, LC, EHL) used an original form to extract data from each full-text article. The original data extraction form was developed to record the study design and the study population (health condition, age). This form enabled broad, general comparisons of the health conditions and demographics of the Para dressage athlete population, and the descriptive and inferential statistics presented on the impairment assessment tool’s reliability and validity.

Reliability was defined as a tool’s stability over time (Streiner, 2015). Included studies evaluated reliability using test-retest, inter, intra-rater and internal consistency, quantified using intraclass correlations (ICC) and kappa statistics. ICC values were evaluated and judged to indicate poor (<0.5), moderate (0.5–0.75), good (0.75–0.9) or excellent (>0.9) reliability (Koo & Li, 2016). Kappa statistics were evaluated and judged to indicate poor (<0.0), slight (0.0–0.2), fair (0.21–0.4), moderate (0.41–0.6), substantial (0.61–0.8) and almost perfect (0.81–1.0) reliability (Landis & Koch, 1977).

Validity was defined as the extent to which a tool measures what it purports to (Streiner, 2015). The included studies used correlation coefficients to examine the strength of associations between the assessment tool and other tools purporting to measure the same construct. Correlation coefficients were evaluated and judged to indicate poor (<0.3), fair (0.3–0.5), moderately strong (0.6–0.8), and very strong (>0.8) associations (Chan, 2003).

A single researcher (LSG) checked the accuracy of data extraction for 10% of the retrieved articles. Responsiveness represents the change in score on a tool that is meaningful to the individual. For example, the change in score of a test between admission and discharge (Gorman et al., 2014a). Responsiveness was not formally assessed in this study but is considered in the discussion section in relation to athletes with fluctuating or progressive impairments (International Paralympic Committee, 2015).

The six authors also completed a bespoke clinical utility score (Appendix B) for each impairment assessment tool that they reviewed. This allowed the practical considerations of athlete classification to be explicitly reflected in the overall appraisal of impairment assessment tools. These factors comprised the test position, the need for and portability of any equipment (as most classification assessments occur at a range of riding venues), the amount of training needed to administer the assessment, the populations it has been validated in, and the type of data produced (ordinal, interval or ratio scaled). Researchers allocated a numerical score for each of seven characteristics (from 0 to 1 or 2; where a lower score indicates greater practicality for Para dressage classification), which were summed to generate an overall usability score (from 0, indicating excellent usability to 11, indicating low usability for Para dressage classification).

A meta-analysis was not appropriate for this study as the methodological diversity across included studies, particularly in relation to the number of measurement and statistical tools that were utilized to quantify reliability, resulted in a substantial level of heterogeneity. Indeed, in one of the few studies to take a meta-analytic approach to reliability (Ehrenbrusthoff et al., 2018), their limitations section confirmed that there was a high amount of heterogeneity and that this aspect of the study should be interpreted with caution. Further, a meta-analysis may obscure important distinctions among the outcomes and may mislead the identification of the most applicable impairment assessment tools. As such, meta-analyses were not conducted in this study.
Results

Identification of potentially suitable impairment assessment tools.

Of 540 impairment assessment tools that were identified from the professional databases, 35 met the suitability criteria for Para dressage. These are presented in Appendix C. From these, 151 keywords were identified and included in the search strategy (see Appendix A).

Literature search to establish the psychometric properties of suitable impairment assessment tools.

We identified 19,109 articles from the systematic search of academic databases, from which 305 articles met the inclusion criteria for data extraction and synthesis. Figure 2 shows a PRISMA diagram summarizing the generation and flow of articles through the synthesis process.

---

**Figure 2.** PRISMA diagram of article flow for the synthesis.
Further refinement of eligible impairment assessment tools based on known performance determinants for Para dressage.

Impairment assessment tools that captured aspects of balance, coordination, muscle tone and strength were included. Using these criteria, the 35 impairment assessment tools were further refined, which resulted in 10 impairment assessment tools. These 10 tools were selected for further examination and data extraction was performed to ascertain their reliability and validity. Range of motion is routinely measured as part of current classification and so was not assessed here. The included impairment assessment tools are presented in Table 1.

Table 1. List of included impairment assessment tools, the impairment they measure, and the key scientific reference for each tool.

<table>
<thead>
<tr>
<th>Selected impairment assessment tools</th>
<th>Impairment measured</th>
<th>Key reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashworth Scale</td>
<td>Spasticity</td>
<td>(Lee et al., 1989)</td>
</tr>
<tr>
<td>Modified Ashworth Scale (MAS)</td>
<td>Spasticity</td>
<td>(Ansari et al., 2008a)</td>
</tr>
<tr>
<td>Function in Sitting Test (FIST)</td>
<td>Sitting balance</td>
<td>(Gorman et al., 2010)</td>
</tr>
<tr>
<td>Scale for Assessment and Rating of Ataxia (SARA)</td>
<td>Coordination</td>
<td>(Schmitz-Hübsch et al., 2006)</td>
</tr>
<tr>
<td>Trunk Impairment Scale (TIS)</td>
<td>Sitting balance/trunk movement</td>
<td>(Verheyden et al., 2004)</td>
</tr>
<tr>
<td>Tardieu Scale</td>
<td>Spasticity</td>
<td>(Paulis et al., 2011)</td>
</tr>
<tr>
<td>Modified Tardieu Scale (MTS)</td>
<td>Spasticity</td>
<td>(Mehrholz et al., 2005a)</td>
</tr>
<tr>
<td>Hand-Held Dynamometry (HHD)</td>
<td>Muscle strength</td>
<td>(Bohannon et al., 2013)</td>
</tr>
<tr>
<td>Motricity Index (MI)</td>
<td>Muscle strength</td>
<td>(Bohannon, 1999)</td>
</tr>
<tr>
<td>Trunk Control Test (TCT)</td>
<td>Trunk movement</td>
<td>(Franchignoni et al., 1997)</td>
</tr>
</tbody>
</table>

Hand-held dynamometry (HHD) and the Motricity Index (MI) for specific muscles of the trunk, shoulder, and hips were included. Hand-held dynamometry (HHD) and the Motricity Index (MI) for specific muscles of the trunk, shoulder, and hips were included because interview data suggested that the strength of these muscle groups is likely to have the greatest impact on Para dressage performance. Further, it is recommended that only the muscle groups that are thought to be influential on sports performance should be tested (Tweedy et al., 2010). The remaining 25 impairment assessment tools were excluded for a range of reasons that are presented in Table 2.

Assessment of reliability, validity, and usability of impairment assessment tools.

We identified two modified versions of impairment assessment tools during data extraction: the Re-Modified Ashworth Scale and Modified Trunk Impairment Scale. Thus, these two tools were included in addition to the 10 impairment assessment tools outlined in Table 1. This resulted in a combined total of 12 impairment assessment tools, as presented in Table 3, alongside details of the populations from which the reliability and validity data were generated and their usability scores. Table 4 shows the validity and reliability of the selected assessment tools. For the Ashworth, Modified Ashworth (MAS), Re-Modified Ashworth scales and Tardieu and Modified Tardieu (MTS) scales, the mean average of either reliability or validity data were used if values for multiple muscles, but not overall scores, were presented.
Table 2. List of excluded impairment assessment tools and the reason(s) for their exclusion.

<table>
<thead>
<tr>
<th>Reason for exclusion of impairment assessment tool</th>
<th>Excluded impairment assessment tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlikely to influence riding performance/could be compensated by adaptations</td>
<td>Jebsen-Taylor Hand Function Test, Sollerman Hand Function Test, Grip Strength, Nine-Hole Peg Test, Purdue Pegboard Test, Toronto Rehabilitation Institute Hand Function Test, Van Lieshout Test Short Version</td>
</tr>
<tr>
<td>Categorization of the severity of a health condition or were only suitable for one condition</td>
<td>National Institutes of Health Stroke Scale, Stroke Rehabilitation Assessment of Movement Measure, Motor Evaluation Scale for Upper Extremity in Stroke, American Spinal Injuries Association’s Impairment Scale, Spinal Cord Assessment Tool for Spastic Reflexes, International Standards for Neurological Classification of Spinal Cord Injury, Fugl Meyer Assessment, Graded Redefined Assessment of Strength, Sensation and Prehension</td>
</tr>
<tr>
<td>Most items performed in standing</td>
<td>Berg Balance Scale</td>
</tr>
<tr>
<td>Already included in classification assessment</td>
<td>Manual Muscle Test, range of movement</td>
</tr>
<tr>
<td>Impractical to test in Para athletes</td>
<td>Biering-Sørensen Test</td>
</tr>
<tr>
<td>Construct already captured on other tools</td>
<td>Scale for Contraversive Pushing and Modified Scale for Contraversive Pushing</td>
</tr>
<tr>
<td>No studies of reliability or validity found</td>
<td>Burke Lateropulsion Scale, Nottingham</td>
</tr>
<tr>
<td></td>
<td>Assessment of Somato-sensations, Grasp and Release Test, Functional Axial Rotation, Catherine Bergogo Scale</td>
</tr>
</tbody>
</table>

The 12 selected impairment assessment tools were assessed across a wide range of health conditions. These are reflective of the conditions and eligible impairments that are common within the population of Para dressage athletes, as recorded by the FEI. As presented in Table 1, these impairment assessment tools provide measures of impairments associated with muscle tone (Ashworth Scale, MAS, Re-Modified Ashworth Scale, Tardieu Scale, MTS), sitting balance (FIST, TIS), coordination (SARA), trunk movement (TIS, Modified TIS, TCT), and strength (Motricity Index, hand-held dynamometry), each of which have been reported to have an impact on dressage performance (Hobbs et al., 2020; St. George et al., 2021). For measures of muscle strength, HHD exhibited the highest collective validity and reliability scores. The standalone tool for measuring coordination, SARA, exhibited excellent collective reliability scores and was deemed to be a valid measurement tool, based on very strong correlations with a range of measures of impairment. Modified versions of the Ashworth and Tardieu scales (MAS, MTS, re-Modified Ashworth Scale) were assessed, and the re-Modified Ashworth scale represented the version with the most robust reliability and validity for assessing muscle tone. Of the tests which assessed sitting balance, trunk movement, or both, the FIST and TIS exhibited superior collective reliability and validity scores and were also deemed the most practical to use in a classification assessment setting, based on usability scores.
Table 3. Details of included scientific articles, populations studied, and the usability score for selected impairment assessment tools.

<table>
<thead>
<tr>
<th>Impairment Assessment Tool</th>
<th>Population (sample size, age, health condition)</th>
<th>Usability Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ashworth Scale</strong></td>
<td>Overall sample size: n = 136 Age (Mean, range) = 54.2, 13 – 81 years Conditions: multiple sclerosis (n = 30), stroke (n = 23), spinal cord injury (n = 24), hemiplegia (n = 59).</td>
<td>2</td>
</tr>
<tr>
<td>(Ansari et al., 2006; Brashear et al., 2002; Fleuren et al., 2010; Haas et al., 1996; Lee et al., 1989; Nakhostin-Ansari et al., 2006; Nuyens et al., 1994; Vattanasilp &amp; Ada, 1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Modified Ashworth Scale</strong></td>
<td>Overall sample size: n = 747 Age (Mean, range) = 53.3, 13 – 90 years Conditions: stroke (n = 278), cerebral palsy (n = 2), multiple sclerosis (n = 1), amyotrophic lateral sclerosis (n = 2), spinal cord injury (n = 163), traumatic brain injury (n = 32), hemiplegia (n = 269).</td>
<td>2</td>
</tr>
<tr>
<td>(Akpinar et al., 2017; Allison et al., 1996; Ansari et al., 2006; Ansari et al., 2008a; Baunsgaard et al., 2016; Blackburn et al., 2002; Bohannon &amp; Smith, 1987; Cooper et al., 2005; Craven &amp; Morris, 2010; Gregson et al., 1999, 2000; Haas et al., 1996; Kaya et al., 2011; Li et al., 2014; Mehrholz et al., 2005a, b; Nakhostin-Ansari et al., 2006; Pandyan et al., 2001, 2003; Sloan et al., 1992; Tederko et al., 2007; Zurawski et al. 2019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Re-Modified Ashworth Scale</strong></td>
<td>Overall sample size: n = 270 Age (Mean, range) = 50.6, 20 - 82 years Conditions: stroke (n = 145), multiple sclerosis (n = 18), spinal cord injury (n = 38), traumatic brain injury (n = 16), hemiplegia (n = 52), tumor (n = 1).</td>
<td>2</td>
</tr>
<tr>
<td>(Ansari et al., 2012; Ansari et al., 2008b, 2009a, b; Ghotbi et al., 2009, 2011; Kaya et al., 2011; Mishra &amp; Ganesh, 2014; Naghdi et al., 2007, 2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FIST</strong></td>
<td>Overall sample size: n = 379 Age (Mean, range) = 56.6, 18 - 94 years Conditions: stroke (n = 243), multiple sclerosis (n = 21), central nervous system neurological condition (n = 7), spinal cord injury (n = 64), traumatic brain injury (n = 12), cancer (n = 12), non-traumatic brain injury (n = 5), deconditioning (n = 2), Guillain-Barre Syndrome (n = 2), hydrocephaly (n = 2), encephalitis (n = 2), chronic inflammatory demyelinating polyneuropathy (n = 2), arteriovenous malformation (n = 1), Parkinson’s Disease (n = 1), Hemicolecctomy (n = 1), medically complex (n = 1), cardiac condition (n = 1).</td>
<td>1</td>
</tr>
<tr>
<td>(Abou et al. 2020; Cabanas-Valdes et al., 2016; Erol et al., 2021; Gorman et al., 2010, 2014a, b; Palermo et al. 2020; Sung et al., 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SARA</strong></td>
<td>Overall sample size: n = 870 Age (Mean, range) = 44.4, 6 - 87 years Conditions: early onset ataxia (n = 38), Friedreich’s ataxia (n = 96), spinocerebellar ataxia (n = 463), non-spinocerebellar ataxia (n = 64), ataxic stroke (n = 60), multiple sclerosis (n = 80), autosomal recessive spastic ataxia of Charlevoix-Saguenay (n = 69)</td>
<td>3</td>
</tr>
<tr>
<td>(Bourcier et al., 2020; Braga-Neto et al., 2010; Brandsma et al., 2017; Bürk et al., 2009; Kim et al., 2014; Salci et al., 2017; Schmitz-Hübsch et al., 2006, 2010; Tan et al., 2013; Weyer et al., 2007; Winser et al., 2018; Yabe et al., 2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impairment Assessment Tool</td>
<td>Population (sample size, age, health condition)</td>
<td>Usability Score</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
</tr>
</tbody>
</table>
| **Tardieu Scale**<sup>(Paulis et al., 2011)</sup> | Overall sample size: n = 13  
Age (Mean, SD): 70.2, 12.3 years  
Condition: stroke (n = 13) | 5 |
| **Modified Tardieu Scale**<sup>(Akpinar et al., 2017; Ansari et al., 2008c; Azarnia et al. 2021; Ben-Shabat et al., 2013; Li et al., 2014; Mehrhozl et al., 2005a; Naghdie et al., 2014, 2017; Santos et al. 2021; Singh et al., 2011; Sonvane & Kumar, 2019)</sup> | Overall sample size: n = 537  
Age (Mean, range): 53.7, 18 - 86 years  
Conditions: stroke (n = 282), multiple sclerosis (n = 56), spinal cord injury (n = 109), hemiplegia (n = 90). | 4 |
| **Trunk Impairment Scale**<sup>(Cabanasa-Valdes et al., 2016; Fil Balkan et al., 2019; Lombardi et al., 2017; Monticone et al., 2017; Parlak Demir & Yildirim, 2018; Quinzanos et al., 2014; Sag et al., 2019; Seo et al., 2008; Verheyden et al., 2004, 2005; 2006a, b; 2007; Zhao et al., 2021)</sup> | Overall sample size: n = 1045  
Age (Mean, range): 60.4, 16 - 95 years  
Conditions: stroke (n = 716), Parkinson’s disease (n = 26), traumatic brain injury (n = 30), multiple sclerosis (n = 30), spinal cord injury (n = 177), myopathy (n = 31), myotonic dystrophy (n = 22), limb girdle muscular dystrophy (n = 7), fascioscapulohumeral muscular dystrophy (n = 3), Becker muscular dystrophy (n = 3). | 1 |
| **Modified Trunk Impairment Scale**<sup>(YunBok et al., 2018)</sup> | Overall sample size: n = 55  
Age (Mean, SD, range): 60.0, 2.5, 57 - 65 years  
Condition: stroke (n = 55). | 1 |
| **Hand-held Dynamometry**<sup>(Aguiar et al., 2016; Akshintala et al., 2021; Aksu & Yakt, 2003; Baschung Pfister et al., 2018; Beck et al., 1999; Bohnmann, 1986, 1992, 1993, 1995; Bohnmann et al., 2013; Bohnmann & Andrews, 1987; Brinkmann, 1994; Busse et al., 2008; Cardin & Bohnmann, 2017; Dyball et al., 2011; Eken et al., 2020; Ekstrand et al., 2015; Faria et al., 2013; Goonetilleke et al., 1994; Hayes et al., 2002; Kilmer et al., 1997; Knak et al., 2020; Larson et al., 2010; Livesley, 1992; May et al., 1997; Mentiplay et al., 2018; Moreno-Navarro et al., 2021; Morris et al., 2008; Noreau & Vachon, 1998; Riddle et al., 1989; Saygin et al., 2020; Schwartz et al., 1992; Tsai et al., 2015; van Langeveld et al., 1996; Visser et al., 2003)</sup> | Overall sample size: n = 1203  
Age (Mean, range): 50.7, 6 - 92 years  
Conditions: stroke (n = 366), spinal cord injury (n = 181), motor neurone disease (n = 38), amyotrophic lateral sclerosis (n = 91), traumatic brain injury (n = 25), myotonic dystrophy type 1 (n = 78), neurofibromatosis type 1 (n = 20), neurofibromatosis type 2 (n = 13), multiple sclerosis (n = 25), dermatomyositis (n = 46), polymyositis (n = 23), necrotizing myopathy (n = 9), anti-synthetase syndrome (n = 11), myositis associated disorder (n = 11), cerebral palsy (n = 73), Guillain-Barre syndrome (n = 1), amputation (n = 1), fracture (n = 1), multiple traumas (n = 1), generalized weakness (n = 1), paraplegia (n = 23), tetraplegia (n = 15), brachial plexus injury (n = 16), hereditary motor and sensor neuropathy type 1 (n = 10), peripheral neuropathy (n = 2), Huntington’s disease (n = 20), orthopedic shoulder issue (n = 17), varied group with weakness (n = 21), non-specified neurological diagnosis (n = 6), neurologic diagnosis other than stroke (n = 14), closed head injury (n = 4), orthopedic diagnosis (n = 5), poliomyelitis anterior acuta (n = 4), Charcot-Marie-Tooth disease (n = 1), Myasthenia Gravis (n = 1), debility (n = 17), radiculopathy (n = 1), spinal muscular atrophy (n = 1). | 3 |
### Table 4. Reliability and validity of included impairment assessment tools.

<table>
<thead>
<tr>
<th>Impairment Assessment Tool</th>
<th>Demographics</th>
<th>Reliability</th>
<th>Validity</th>
<th>Overall rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ashworth Scale</strong></td>
<td>Total number of studies: 8, Total number of participants: 136</td>
<td>Interrater: ICC = 0.58, Kendall's Coefficient of Concordance = 0.92, Kendall's Tau = 0.39, Cohen's Weighted Kappa = 0.2</td>
<td>SRCC: EMG = 0.39</td>
<td>Reliability: Slight to substantial/moderate. Validity: Fair</td>
</tr>
<tr>
<td>(Ansari et al., 2006; Brashear et al., 2002; Fleuren et al., 2010; Haas et al., 1996; Lee et al., 1989; Nakhhostin-Ansari et al., 2006; Nuyens et al., 1994; Vattanasilp &amp; Ada, 1999)</td>
<td>Average age: 54.2</td>
<td>In Intrarater: Coefficient of Repeatability = 8, Coefficient of Variation = 6.4, Cohen's Weighted Kappa = 0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Modified Ashworth Scale</strong></td>
<td>Total number of studies: 22, Total number of participants: 747</td>
<td>Interrater: ICC = 0.56, SRCC = 0.63, Kendall's Tau = 0.63, Cohen's Weighted Kappa = 0.32</td>
<td>SRCC: EMG magnitude = 0.21</td>
<td>Reliability: Fair to moderate. Validity: Poor</td>
</tr>
<tr>
<td>(Akpinar et al., 2017; Allison et al., 1996; Ansari et al., 2006; Ansari et al., 2008a; Baunsgaard et al., 2016; Blackburn et al., 2002; Bohannon &amp; Smith, 1987; Cooper et al., 2003; Craven &amp; Morris, 2010; Gregson et al., 1999, 2000; Haas et al., 1996; Kaya et al., 2011; Li et al., 2014; Mehrholz et al., 2005a,b; Nakhhostin-Ansari et al., 2006; Pandyan et al., 2001, 2003; Sloan et al., 2008)</td>
<td>Average age: 53.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impairment Assessment Tool</td>
<td>Demographics</td>
<td>Reliability</td>
<td>Validity</td>
<td>Overall rating*</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Re-Modified Ashworth Scale</strong> (Ansari et al., 2012; Ansari et al., 2008b, 2009a, b; Ghotbi et al., 2009, 2011; Kaya et al., 2011; Mishra &amp; Ganesh, 2014; Naghdi et al., 2007, 2008)</td>
<td>Total number of studies: 10, Total number of participants: 270 Average age: 50.6</td>
<td>Interrater: Kendall's Tau-b = 0.85, Cohen's Weighted Kappa = 0.82, Intrarater: Kendall's Tau-b = 0.92, Cohen's Weighted Kappa = 0.82</td>
<td>SRCC: H-indexes = 0.58</td>
<td>Reliability: Almost perfect Validity: Moderately strong</td>
</tr>
<tr>
<td><strong>FIST</strong> (Abou et al., 2020; Cabanas-Valdes et al., 2016; Erol et al., 2021; Gorman et al., 2010, 2014a, b; Palermo et al., 2020; Sung et al., 2016)</td>
<td>Total number of studies: 8, Total number of participants: 376 Average age: 56.6</td>
<td>Interrater: ICC = 0.98 Intrarater: ICC = 0.98 Test-retest: ICC = 0.95 Consistency: Cronbach's Alpha = 0.93</td>
<td>SRCC: BBS = 0.81, Dynamic Balance Grade = 0.93, Sitting Balance Grade = 0.93, TIS = 0.80, FIM = 0.81 PCC: VTC = 0.23, mFRT (lateral) = 0.64, mFRT (forward) = 0.16</td>
<td>Reliability: Excellent Validity: Poor to very strong</td>
</tr>
<tr>
<td><strong>SARA</strong> (Bourcier et al., 2020; Braga-Neto et al., 2010; Brandsma et al., 2017; Bürk et al., 2009; Kim et al., 2014; Salci et al., 2017; Schmitz-Hübsch et al., 2006, 2010; Tan et al., 2013; Weyer et al., 2007; Winser et al., 2018; Yabe et al., 2008)</td>
<td>Total number of studies: 12, Total number of participants: 870 Average age: 44.4</td>
<td>Interrater: ICC = 0.98, PCC = 0.86 Intrarater: ICC = 0.97 Test-retest: ICC = 0.98 Consistency: Cronbach's Alpha = 0.90</td>
<td>SRCC: BARS = 0.94, ICARS = 0.80, 10mWT = -0.77, SFNT = -0.82, TUG = 0.81, 30-CST = 0.92, BBS = 0.95, BI = -0.91, PG-ICARS = 0.92 PCC: FARS = 0.94, ICARS = 0.78, DSI-ARSACS = 0.95, LEMOCOT = -0.87</td>
<td>Reliability: Excellent Validity: Moderate to very strong</td>
</tr>
<tr>
<td><strong>Tardieu Scale</strong> (Paulis et al., 2011)</td>
<td>Total number of studies: 1, Total number of participants: 13 Average age: 70.2</td>
<td>Interrater: ICC = 0.75 Test-retest: ICC = 0.84</td>
<td></td>
<td>Reliability: Good Validity: NA</td>
</tr>
</tbody>
</table>
### Impairment Assessment Tool

<table>
<thead>
<tr>
<th>Tool</th>
<th>Demographics</th>
<th>Reliability</th>
<th>Validity</th>
<th>Overall rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modified Tardieu Scale</strong></td>
<td>Total number of studies: 12, total number of participants: 537, average age: 53.7</td>
<td>Interrater: ICC = 0.73, Cohen's Weighted Kappa = 0.50, Intrarater: ICC = 0.68, Cohen's Weighted Kappa = 0.74, Test-retest: ICC = 0.84, Correlation (unspecified) = 0.94, Consistency: SRCC = 0.95</td>
<td>SRCC: ( \text{Hmax/Mmax} = -0.04 ), Hslp/Mslp = 0.24, MAS = 0.81</td>
<td>Reliability: Moderate to substantial/good, Validity: Poor to very strong</td>
</tr>
<tr>
<td><strong>Trunk Impairment Scale</strong></td>
<td>Total number of studies: 14, total number of participants: 1045, average age: 60.4</td>
<td>Interrater: ICC = 0.97, Cohen's Weighted Kappa = 0.99, Cronbach's alpha = 0.996, Intrarater: ICC = 0.97, Cronbach's alpha = 0.995, Test-retest: ICC = 0.95, Cohen's Weighted Kappa = 0.999, Cronbach's alpha = 0.98, SRCC = 0.97, Consistency: Cronbach's Alpha = 0.88</td>
<td>SRCC: TCT = 0.80, BBS = 0.81, MBI = 0.84, PCC: SCIM = 0.87, BI = 0.78, BBS = 0.89, RMI = 0.78, SF-36 = 0.60</td>
<td>Reliability: Excellent/almost perfect, Validity: Moderately strong to very strong</td>
</tr>
<tr>
<td><strong>Modified Trunk Impairment Scale</strong></td>
<td>Total number of studies: 1, total number of participants: 55, average age: 60.0</td>
<td></td>
<td>SRCC: BBS = 0.82, Postural Assessment Scale = 0.55, TCT = 0.63</td>
<td>Reliability: NA, Validity: Moderately strong to very strong</td>
</tr>
<tr>
<td><strong>Hand-held Dynamometry</strong></td>
<td>Total number of studies: 35, total number of participants: 1203, average age: 50.7</td>
<td>Interrater: ICC = 0.87, PCC = 0.98, SEM = 1.1, Intrarater: ICC = 0.94, PCC = 0.94, SEM = 2.35, SRCC = 0.85, Coefficient of Variation = 0.99, SEM = 0.74, ANOVA = 1.66, Test-retest: ICC = 0.92, PCC = 0.93, SEM = 0.04</td>
<td>SRCC: MI = 0.96, TUG = -0.32, PCC: Isokinetic Dynamometry = 0.9, MVIC = 0.87, TUG = -0.34, STS = 0.11, MMT = 0.48, 10mWT = 0.49</td>
<td>Reliability: Good to excellent/very strong, Validity: Poor to very strong</td>
</tr>
<tr>
<td>Impairment Assessment Tool</td>
<td>Demographics</td>
<td>Reliability</td>
<td>Validity</td>
<td>Overall rating*</td>
</tr>
<tr>
<td>----------------------------------------------------------------</td>
<td>------------------------------</td>
<td>--------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Hayes et al., 2002; Kilmer et al., 1997; Knak et al., 2020; Larson et al., 2010; Livesley, 1992; May et al., 1997; Mentiplay et al., 2018; Moreno-Navarro et al., 2021; Morris et al., 2008; Noreau &amp; Vachon, 1998; Riddle et al., 1989; Saygin et al., 2020; Schwartz et al., 1992; Tsai et al., 2015; van Langeveld et al., 1996; Visser et al., 2003)</td>
<td>Inter-session: ICC = 0.94, PCC = 0.96, SEM = 0.10</td>
<td>Parallel Forms Reliability: PCC = 0.82, Consistency: Cronbach's Alpha = 0.95</td>
<td>SRCC: RMA = 0.77, PCC: Dynamometry Value = 0.78</td>
<td>Reliability: Very strong. Validity: Moderately strong</td>
</tr>
<tr>
<td>Motricity Index (Bohannon, 1995, 1999; Cameron &amp; Bohannon, 2000; Collin &amp; Wade, 1990; Fayazi et al., 2012; Vos-Vromans et al., 2005)</td>
<td>Total number of studies: 6, Total number of participants: 110 Average age: 60.3</td>
<td>Interrater: SRCC = 0.88 Test-retest: SEM = 4.7</td>
<td>SRCC: RMA = 0.75, FIM = 0.50, TIS = 0.68</td>
<td>Reliability: Moderately strong to excellent Validity: Fair to moderately strong</td>
</tr>
<tr>
<td>Trunk Control Test (Collin &amp; Wade, 1990; Fil Balkan et al., 2019; Franchignoni et al., 1997; Parlak Demir &amp; Yildirim, 2015)</td>
<td>Total number of studies: 4, Total number of participants: 222 Average age: 55.2</td>
<td>Interrater: SRCC = 0.76 Intradater: ICC = 0.98 Test-retest: SRCC = 0.96 Consistency: Cronbach's Alpha = 0.81</td>
<td>SRCC: RMA = 0.75, FIM = 0.50, TIS = 0.68</td>
<td>Reliability: Moderately strong to excellent Validity: Fair to moderately strong</td>
</tr>
</tbody>
</table>

Key: MI – Motricity Index; MVIC – Maximal Voluntary Isometric Contraction; RMA and RMI – Rivermead Mobility Assessment/Index; BBS- Berg Balance Scale; TCT – Trunk Control Test; TIS – Trunk Impairment Scale; SCIM – Spinal Cord Independence Measure; EMG – Electromyography; SRCC – Spearman’s Rank Correlation Coefficient; PCC – Pearson’s correlation coefficient. ICC-Intra class correlation coefficient; SEM – Standard Error of the Mean; FIM – Functional Independence Measure; VTC – virtual time to contact; mFRT – modified functional reach test; 10mWT – 10 meter walk test; DSI-ARSACS – Disease Severity Index for Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay; LEMOCOT – Lower Extremity Motor Coordination Test; SNFT – Standardized Finger Nose Test; TUG – Timed Up and Go Test; 30-CST – 30 Second Chair Stand Test; BI – Barthel Index; MBI – Modified Barthel Index; ICARS – International Cooperative Ataxia Rating Scale; PG-ICARS – Posture and Gait sub-component of ICARS; MMT – Manual Muscle Test; MAS – Modified Ashworth Scale; STS – Sit-to-Stand Test; SF-36 – Evaluation of General Quality of Life : Short form-36. *Judgement of overall reliability and validity based upon published criteria (Chan, 2003; Koo & Li, 2016; Landis & Koch, 1977).
Classification of Para athletes, based on sport-specific empirical evidence, is vital to ensure equitable competition. As part of a wider research project to develop a strong evidence base for Para dressage classification, identification of existing impairment assessment tools was necessary for future assessments of the association between impairment and activity limitation. By articulating our search terms, processes, and supporting evidence that underpinned the decisions regarding potential classification tools, we aimed to present a clear and reproducible approach for conducting a synthesis of existing literature that has both scientific rigor and practical usefulness for Para sport classification.

The multi-stage approach used in this synthesis is a transparent method to select existing impairment assessment tools for Para sport classification. Each stage was distinctly different. In the first stage we screened two online, professional databases of rehabilitation measures to identify impairment assessment tools that quantitatively measured eligible impairments for Para dressage in seated, supine, or prone positions. In stage two, we conducted a scientific literature search, using a search strategy developed from the professional database screening results. In stage three, we refined selected impairment assessment tools based on known performance determinants for Para dressage. Finally, we assessed and refined the remaining impairment assessment tools based on their reported reliability and validity, and by scoring their practical usability for Para dressage classification purposes. This approach led to the identification of impairment assessment tools that had acceptable psychometric properties, captured the physical requirements for Para dressage performance (Hobbs et al., 2020; St George et al., 2021), and could meet the practical demands of athlete classification. In our view, the approach represents an important addition to the work of others who have sought to ensure that Para athlete classification systems are transparent and parsimonious (Connick et al., 2018a; Tweedy et al., 2010, 2014).

The synthesis identified 12 impairment assessment tools that were deemed suitable for assessing the impairments of Para dressage athletes. The selected impairment assessment tools provide measurements of muscle tone, followed by trunk impairment and sitting function, muscle strength, and finally, coordination, each of which are recognized to be important determinants of Para dressage performance (Hobbs et al., 2020; St. George et al., 2021). From these 12 tools, further refinement identified the SARA (coordination), Remodified MAS (muscle tone), HHD (muscle strength), FIST and TIS (sitting balance, trunk movement, or both) to be superior in reliability, validity, or both for each impairment that they assess.

When assessing the suitability of identified impairment assessment tools, one must consider sport-specific performance demands, which best represent the fundamental skills, abilities, and physical positions required for sport-specific performance (Tweedy et al., 2016). Thus, for Para dressage athletes, body positioning during impairment assessment must reflect performance demands during riding, as described by Byström et al. (2015) and Eckhardt and Witte (2016), amongst others. Notably, these performance demands include the seated riding position and, where applicable, upper peripheral limb position for rein carrying. As such, we included only impairment assessment tools that were conducted in seated, supine, or prone positions. Impairment assessment tools that employ a seated position have previously been used for athletes with spinal column injury (Roldan et al., 2020) and are considered appropriate for the classification of Para athletes (Altmann et al., 2016; Larson, 2010; Nicholson et al., 2018; Smith et al., 2021). This builds upon recent considerations by Smith and colleagues (2021), who highlighted that the validity of impairment assessment tools for classification may be threatened when practical limitations of the test application are recognized. As such, we recommend that future studies of Para
dressage athletes should incorporate impairment assessments that employ a seated position and should also avoid the measurement of joints or movements that are not applicable to the performance demands of Para dressage.

The FIST and TIS are performed in a seated position and were therefore deemed appropriate for use on Para dressage athletes. For activity-based deficits and performance relating to sitting balance, the FIST was identified as a reliable measure (Gorman et al., 2014b) as was the TIS (Verheyden et al., 2004) for assessing trunk control (Dasoju et al., 2021). Trunk impairment has been assessed in wheelchair rugby athletes and greater trunk impairment severity was associated with reduced sports performance, as quantified through the ability to perform chair tilt and acceleration in the first 2 metres of movement (Altmann et al, 2018). The importance of assessing trunk impairment is justified when considering the wider context of performance determinants for Para dressage (Hobbs et al., 2020; St. George et al., 2021), particularly given the unique requirement for athletes to maintain dynamic posture and remain “in harmony” with the horse’s movement (Peham et al., 2001; Olivier et al., 2017; St. George et al., 2021). That said, repetition of elements within the FIST and TIS should be acknowledged and avoided when developing any impairment assessment protocols for the classification of Para dressage athletes. Omitting repetition or redundant test items across similar impairment assessment tools is imperative, as it may reduce the level of fatigue experienced by an athlete during physical assessments and should therefore be assessed in future studies.

The most universally accepted tool for the quantification of muscle tone is the MAS (Meseguer-Henarejos et al., 2018; Harb & Krishner, 2022), which is a modified version of the Ashworth Scale. Although it has been subject to some criticism, mainly due to its limited ability to distinguish between factors contributing to passive stretch resistance (Harb & Kishner, 2022), it is quick to perform and has no instrumentation requirements (Craven & Morris, 2010). These traits are beneficial in relation to the practical usability of the tool for classification purposes. The MAS was re-modified by Ansari et al. (2006) and has since been reported to have improved reliability and validity. There was superior reliability of the Re-Modified MAS, especially when compared to the other measures of muscle tone. Taken together, the practical usability and superior reliability of the Re-modified MAS supports its potential use for measuring muscle tone in Para dressage athletes.

The SARA tool assess ataxia. It has excellent reliability and is thus a useful measure for quantifying impaired coordination of voluntary muscle movement. Evidently, the elements that make up this robust impairment assessment tool may compliment the assessment of physical impairments caused by ataxia, which is present within the population of Para dressage athletes. Rider coordination makes a significant contribution to overall dressage performance (Peham et al., 2001, Hobbs et al., 2020; St. George et al., 2021), yet the relationship between measures of ataxia and functional performance has not yet been established within the sport of Para dressage. Further, SARA does include test items related to ambulation and speech, which may not be relevant for Para dressage performance. Thus, it would likely be necessary to omit the ambulation and speech items from the SARA tool when assessing athlete impairment in relation to Para dressage performance. That said, an adapted version of SARA has the potential to make an important contribution to the quantification of lower and upper peripheral limb coordination impairments in Para dressage athletes.

Muscle strength is often quantified for athlete classification across a range of Para sports (Paix et al., 2021). In these instances, accurate assessment of maximal voluntary contraction (MVC) is essential, yet maximal voluntary effort (MVE) is reliant on participant cooperation. Consequently, submaximal efforts are likely to occur where the potential for personal gain is favourable, particularly when it may influence sport class allocation (Paix et al., 2021).
HHD is a common method of quantifying isometric muscle strength (Bakers et al., 2021) and the association between muscle strength and performance in Para sport populations has been studied (Smith et al., 2021). In Para wheelchair athletes, greater deficits in performance were observed in those with more severe impairment in trunk strength (Altmann et al., 2016; 2018; Roldan et al., 2020). Further, lower isometric strength scores of the upper peripheral limb were reported for Para swimmers with physical impairments, compared to non-disabled participants, suggesting that the application of HHD is useful for the quantification in Para swimming populations (Hogarth et al., 2019). However, it is difficult to extrapolate these findings to Para dressage, as performance is related to the coordination and grading of muscular contraction for instructing to the horse, or to maintain dynamic postural control (St. George et al., 2021). Thus, future research is required to assess the strength of association between muscle strength and Para dressage performance. HHD could be beneficial for this future research, as it permits quantification of both mean strength values, which may represent important muscular functionality for Para dressage (Simpson, 2019; St. George et al., 2021), as well as MVCs, which may also be relevant for dressage performance (Hobbs et al., 2020). Taken together, we suggest that HHD represents an appropriate impairment assessment tool for Para dressage athletes, particularly given the fact that it produces ratio scaled data.

The judicious selection of existing impairment assessment tools, that are currently and widely used in clinical populations, brought several strengths to this study. These include selecting robust impairment assessment tools that have been rigorously tested in relation to their reliability and validity, which has been established across a range of clinical conditions. It should be noted that the judgement of validity, based on the strength of associations between two tools that are thought to measure similar constructs, was made by the lead researcher (RS) who is a physiotherapist and qualified to make these decisions. For some of the selected outcome tools, the associations between similar tools were obvious and straightforward for example, the HHD and fixed dynamometry both measure strength in a similar way (May et al., 1997). For others however, (e.g., Ashworth and Tardieu tools), the tools used for comparison of validity, such as, electromyography (Fleuren et al., 2010), H-Reflexes (Naghdi et al., 2014) do not necessarily measure the same construct. This may mean that statistical associations are markedly weaker when compared to other tools, as the nature of the comparison tool is considered alongside the strength of its associations with the assessment tool when judging validity and responsiveness. Responsiveness of a tool is important in relation to Para athletes, as their physical conditions may deteriorate over time and their classification status may require review. In previous research, a clinically meaningful change is indicated for the FIST (Gorman et al., 2014a) and HHD is reported as being capable of detecting changes in stroke patients, but with limited responsiveness (Bohannon et al., 2013). The Italian version of the TIS was reported to be responsive in acute and chronic stroke patients (Monticone et al., 2017), and SARA performed best along with a visual analogue scale for classifying worsening spinocerebellar ataxia patient condition (Schmitz-Hübsch et al., 2010). These results suggest that such impairment assessment tools may be useful for detecting changes in Para athletes with a review status, but further research is required to confirm this.

HHD is the only impairment assessment tool identified in this synthesis which produces ratio scaled data, as recommended by the IPC (International Paralympic Committee, 2015). All other identified impairment assessment tools produce ordinal data, which can contribute to poor reliability (Beckman et al., 2017; Van De Pol et al., 2010) and can be difficult to assess in relation to performance, as has been acknowledged in Para swimming (Smith et al., 2021). In coming to a total score for these ordinal scales, data are treated as intervals, but this does not mean that they have all the characteristics of parametric data. For example, the presence of a tool's classification status may influence the use of that tool.
of a ceiling effect is unlikely for a ratio level assessment tool, but may be evident for some of the selected ordinal scales. A ceiling effect is particularly pertinent, because most assessment tools seek to assess proficiency, yet Para sport classification assessment tools need to reflect excellence. As a consequence, Para athletes may outperform others drawn from clinical populations. It is also evident that some tools (notably those assessing muscle tone) only seek the presence and severity of a single form of an impairment, (i.e., hypertonia; Ganguly et al., 2021) and would not reflect the hypotonia present in people with lower motor neuron disorders (Garg et al., 2017). Although this might suggest that muscle tone measurement should not be included in Para athlete assessments, others have reported that the presence of hypertonia can obfuscate strength measurements in Para athletes (Hogarth et al., 2019). Further, St. George and colleagues (2021) reported the perceived importance of muscle tone to Para dressage, necessitating its measurement.

An additional and important point to consider when contextualizing the findings of this synthesis is the assumption that impairment-based measurements can accurately predict activity limitations, in this case, sports performance. This is at odds with wider disability literature and the WHO International Classification of Functioning, Disability and Health, which stipulates that those with similar impairments do not necessarily produce comparable activity limitations (World Health Organization, 2001). For example, there are only weak associations between impairments and activity in people with neuromuscular disease (Vandervelde et al., 2009). Others have reported an absence of strong associations between impairment-based classifications and an individual’s performance in Para sport (Burkett et al., 2018). This may be because individuals develop compensatory strategies and have different personal motivational and contextual factors, such as, access to aids, which all enable individuals to perform at a different level than their impairments might predict (Vandervelde et al., 2009).

Collectively, findings from this synthesis underline the importance of rigorously testing the performance of selected impairment assessment tools in future work. This is necessary to advance the development of an evidenced-based classification system for Para dressage, which thoroughly considers the performance demands of the Para dressage athlete. The design of future studies would be underpinned by the results from this synthesis. In accordance with recent recommendations (Paix et al., 2021), an upcoming study will test the performance of the selected impairment assessment tools alongside existing classification methods in Para dressage athletes. It will be important to ensure that all test items, as well as summary scores, from each ordinal impairment assessment tool are recorded to provide a granular understanding of which items provide valuable data for classification and which ones do not, rather than relying on a total score, to overcome any potential ceiling effect. These findings should be statistically compared with sports-specific assessments, in this case, taken during standardized simulated riding. Detailed statistical analysis, for example using cluster analysis (Connick et al., 2018b) will be necessary to determine the validity of the impairment assessment tools and to evaluate which ones are most associated with Para dressage performance. Inclusion of able-bodied athletes in future research is essential to provide a range of normative data and to inform decisions regarding the minimal impairment criteria. The introduction of any new classification tools and criteria require thoughtful implementation to ensure that classifiers receive, and can access, thorough training. The inter-rater reliability of classifiers using new impairment assessment tools for classification purposes could also be rigorously evaluated and regular fidelity checks undertaken to ensure that Para dressage athletes are assessed consistently.

Limitations
There are some limitations to this work. Whilst the assessment of usability was important to ensure that impairment assessment tools could be used ‘in the field’, the bespoke usability scale was developed for this study and therefore lacks any external assessment of its validity. However, the criteria used (presented in Appendix C) was modelled after the clinical utility scoring system that was used in the systematic reviews of Tyson and Connell (2009a, b) and developed with current FEI classifiers which confers some external validity. Inclusion of interview data (St. George et al., 2021) to guide selection of impairment assessment tools may also be prone to some bias. That said, we and others argue that the views of stakeholders and the usability of the tools ‘in the field’ are important to their uptake and usefulness (Connick et al., 2018a; St. George et al., 2021).

The pooling of validity and reliability estimates from a range of studies in different populations with health conditions will obviously add variability to the findings but does increase the applicability of the results to the diverse group of Para athletes that participate in Para dressage. There is also likely to be a bias towards publishing positive reports of the validity and reliability of impairment assessment tools which should be considered when viewing the results. However, as all of the included impairment assessment tools are free to use, there is perhaps less likelihood that commercial interests influence reporting in the scientific studies included here.

A potential limitation of the literature review is that, whilst conducted using a systematic approach, it would not conform to systematic review guidelines utilized by groups such as the Cochrane Collaboration (Higgins et al., 2019). However, the use of these guidelines would not enable the aims of this study to be met, particularly as it would not enable use of stakeholder views and assessment of practical usability to refine selection. Further, it is unlikely that randomized trials, comparing assessment tools within the broad range of populations, would be identified if these guidelines were employed in this study. Consequently, it would not result in the selection of tools that were useful for Para dressage.

Conclusions

This synthesis identified 12 impairment assessment tools that could be used to measure the eligible impairments of Para dressage athletes and could thus be used in future research on the development of recommendations for an evidence-based Para dressage classification system. The strength of using existing impairment assessment tools in this way provides rapid access to substantial data evaluating their validity and reliability across a range of conditions, which are present within the population of Para dressage athletes. Robust data on the validity and reliability of new impairment assessment tools would take many years to accrue, limiting judgements on their validity and usefulness. The limitations of using existing impairment assessment tools to classify Para dressage athletes does however, include the potential presence of a ceiling effect on some tools, an unknown relationship to sports performance, and the predominance of tools that yield ordinal, rather than ratio scaled data. Further research should investigate the performance of these existing impairment assessment tools within the Para dressage population. Further research should also seek to establish the inter-rater reliability of these tools for classification purposes to ensure that that Para athletes and the wider Para sport community can have confidence in the measurements.
Perspectives

Accurate and transparent Para sport classification means that athletes can compete fairly against others (Connick et al., 2018b; Tweedy et al., 2014). This synthesis presents a novel process by which impairment assessment tools were selected, refined, and critically examined. Existing knowledge of the determinants of dressage performance (Hobbs et al., 2020), the views and experiences of Para dressage stakeholders (St George et al., 2021), analysis of the validity and reliability of eligible impairment assessment tools across a range of conditions, and the assessment of their practical utility within a classification setting were used. The methods and processes outlined and employed here could be readily adapted for other Para sports to enable them to identify existing impairment assessment tools that are reliable, valid, and practically feasible classification assessment tools.

Author affiliations:
1 Faculty of Community Health, University of Central Lancashire, Preston UK PR1 2HE; RStockley1@uclan.ac.uk, JSpencer11@uclan.ac.uk
2 Research Centre for Applied Sport, Physical Activity and Performance, University of Central Lancashire, Preston UK PR1 2HE; LStgeorges@uclan.ac.uk, SJHobbs1@uclan.ac.uk, JAlexander3@uclan.ac.uk
* Correspondence: RStockley1@uclan.ac.uk; Tel.: +44-1772-804998

Author Contributions: Conceptualization, RCS, LSG, SJH; Methodology, RCS, LSG, SJH; Formal Analysis, LSG, JS; Writing-Original Draft Preparation, RCS, LSG; Writing-Review & Editing, RCS, LSG, JS, SJH, JA.

Funding: This work was funded by the Fédération Équestre Internationale, Switzerland FEI Para-Equestrian sport: 2018–2021.

Acknowledgments: E.H. Littlefair, L. Cross and P. Stainton for additional screening and data extractions. J. Reed for running the search strategy and C. Dickinson for co-development of the search strategy and the initial professional database screening. S. Gregory is acknowledged for advice regarding classification and tool selection.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References


Neurosurgery & Psychiatry, 88(6), 474-483. [http://dx.doi.org/10.1136/jnnp-2016-313526](http://dx.doi.org/10.1136/jnnp-2016-313526)


Appendix A

Search Strategy

Impairment

1. Deterioration
2. spasticity
3. hypertonia
4. contracture
5. tetraplegia
6. Hemiparesis
7. Hemiplegia
8. Impaired muscle power
9. Passive range of movement
10. Limb deficiency/ inequality
11. Leg length difference/inequality
12. Short stature
13. Ataxia
14. Athetosis
15. Somatosensory Impairment
16. Quadriplegia
17. fatigue
18. Motor deficit
19. paresis
20. sensory impairment
21. paraplegia
22. motor disorder"
23. "motor impairment"
24. OR/1-23

Measures

25. muscle
26. contraction
27. strength
28. power
29. tone/tonus
30. function
31. anthropometry
32. Neuromuscular
33. Postur*
34. endurance
35. control
36. Motricity
37. Isometric
38. Balance
39. Coordination
40. motor
41. Electromyography
42. Goniometry
43. inertial sensor/ inertial measurement unit/ IMU
44. biomechanic*
45. dexterity
46. equilibrium
47. movement
48. kinematic*
49. "Joint Range of Motion" / "Range of Motion" / "ROM"
50. force
51. sensation
52. "Proportional control"
53. "hand strength"
54. "hand dexterity"
55. "Postur* control"
56. "Postur* balance"
57. "muscle activity"
58. "muscle contraction"
59. "Muscle endurance"
60. "muscle function*"
61. "muscle strength"
62. "muscle power"
63. "Muscle tone/ tonus"
64. "Muscle weakness"
65. "motor activity"
66. "motor coordination"
67. "motor skill*"
68. "motor function"
69. "motor evaluation"
70. "motor recovery"
71. "motricity index"
72. "Functional decline"
73. "Force control"
74. "physical endurance"
75. "isometric strength"
76. "isometric contraction"
77. "neuromuscular control"
78. OR/25-77
Assessment
79. Ashworth Scale
80. Modified Ashworth Scale
81. Jebsen-Taylor Hand Function Test
82. Tardieu Scale
83. Modified Tardieu Scale
84. International Standards for Neurological Classification of Spinal Cord Injury
85. ASIA Impairment Scale
86. Burke Lateropulsion scale
87. Toronto Rehabilitation Institute Hand Function Test
88. Purdue pegboard test
89. Scale for contraversive pushing
90. Dynamometer/ Dynamometry
91. Sollerman Hand Function Test
92. Function in Sitting Test
93. Trunk Impairment Scale
94. Nine-Hole Peg Test
95. Modified SCP
96. Modified Scale for Contraversive Pushing
97. Motor Evaluation Scale for Upper Extremity in Stroke
98. Spinal Cord Assessment Tool for Spastic Reflexes
99. Graded and Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP)
100. Manual Muscle Test
101. Grasp and Release Test
102. Hand Held Myometer / Myometry
103. Hand Held Dynamometer / Dynamometry
104. Hand-held Grip Strength
105. Motricity Index
106. Functional Axial Rotation
107. Bierno-Sorensen Test
108. Catherine Bergogo Scale
109. Scale for assessment and rating of ataxia (SARA)
110. National Institutes of Health Stroke Scale
111. NIH Stroke Scale
112. Berg Balance Scale
113. Fugl-Meyer Assessment of Motor Recovery
114. Stroke Rehabilitation Assessment of Movement Measure
115. Jebson Hand Function Test
116. Trunk Control Test
117. Nottingham Assessment of Somatosensations
118. Van Lieshout Test Short Version
119. OR/79-118

General Terms
120. evaluation
121. reliability
122. validity
123. scale
124. Impair*
125. Reproducibility
126. Responsiveness
127. disability
128. Function*
129. Rehabilitation
130. Diagnostic*
131. Psychometric*
132. Assessment*
133. Outcome*
134. Measure*
135. Tool
136. instrument
137. pain
138. physiotherapy
139. sensitivity
140. specificity
141. examination
142. test

143. clinimetric*
144. test-retest
145. "concurrent validity"
146. "construct validity"
147. "inter-rater reliability"
148. "intra-rater reliability"
149. "observer variation"
150. "Reference value"
151. "Reproducibility of Results"
152. "Body structure"
153. "body function"
154. "occupational therapy"
155. "occupational therapy evaluation"
156. "physical therapy"
157. "Physical examination"
158. "Outcome* measure"
159. "Outcome* assessment"
160. "outcome* tool*"
161. "pain assessment"
Appendix B

<table>
<thead>
<tr>
<th>Usability Rating Scale</th>
<th>FEI Requirement</th>
<th>Rating</th>
<th>Score</th>
<th>Record Info/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td>Sitting, supine or prone lying</td>
<td>Sitting = 0 Supine or prone lying or sitting = 1 Standing or ambulation required = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment Required</td>
<td>Low cost, portable, user-friendly equipment. Or no equipment</td>
<td>No equipment = 0 If 0 scored here, score the two rows below with “n/a” User-friendly = 0 Training required, but relatively user-friendly = 1 Highly-specialised skills required = 2 Portable = 0 Large or heavy, but portable = 1 Not portable (stationary) = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required Training</td>
<td>1 – 2 days training for new impairment measures, which may include online training</td>
<td>Simple training (i.e. paper manual, standardised printed instructions) = 0 Moderate training (i.e. online video instructions or course) = 1 One-to-one training = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient population</td>
<td>Generic</td>
<td>Condition specific (only one patient group tested, excluding control group) = 1, Generic (more than one patient group tested) = 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data produced</td>
<td>Interval or ratio scaled</td>
<td>Interval or Ratio scaled = 0 Ordinal = 1 No quantitative data produced = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td></td>
<td></td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

FEI – Fédération Équestre Internationale
Appendix C

Potentially eligible impairment tools

1. Ashworth Scale
2. Function in Sitting Test
3. Scale for Assessment and Rating of ataxia (SARA)
4. Trunk Impairment Scale
5. Tardieu Scale
6. Hand-held Dynamometry
7. Motricity Index
8. Trunk Control Test
10. Nine-Hole Peg Test
11. Grip Strength
12. Stroke Rehabilitation Assessment of Movement Measure
13. Scale for Contraversive Pushing
14. Modified Scale for Contraversive Pushing
15. Biering-Sørensen Test
16. Manual Muscle Test
17. Jebsen-Taylor Hand Function Test
18. Sollerman Hand Function Test
19. National Institutes of Health Stroke Scale
20. Spinal Cord Assessment Tool for Spastic Reflexes
22. Fugl-Meyer Assessment of Motor Recovery
23. Graded and Redefined Assessment of Strength, Sensibility, and Prehension (GRASSP)
24. ASIA Impairment Scale
25. Toronto Rehabilitation Institute Hand Function Test
26. Motor Evaluation Scale for Upper Extremity in Stroke
27. Purdue Pegboard Test
28. Functional Axial Rotation
29. Burke Lateropulsion Scale
30. Nottingham Assessment of Somato-sensations
31. Grasp and Release Test
32. Modified Tardieu Scale
33. Catherine Bergogo Scale
34. Van Lieshout Test Short Version
35. Modified Ashworth Scale.

© 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).