EUJAPA

Review

A synthesis of potential impairment assessment tools for Para dressage classification

Rachel C Stockley[1], Lindsay St George[2], Jill Alexander[2], Joseph Spencer[1], and Sarah Jane Hobbs[2]

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.

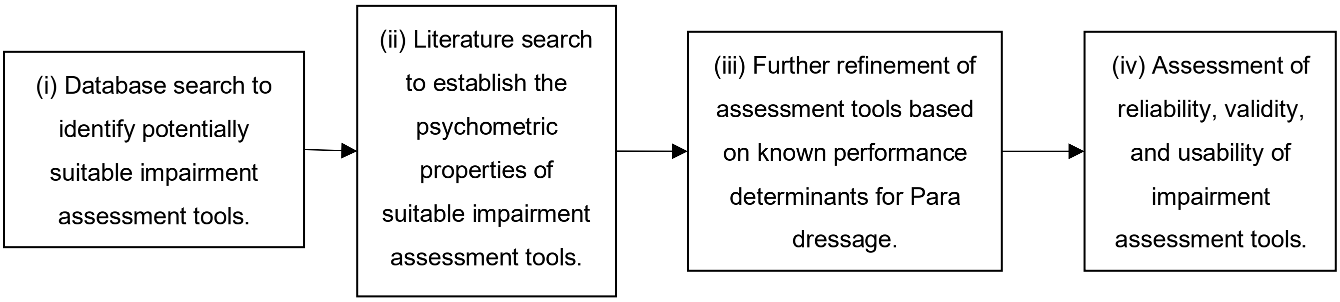

Figure 1. Stages of the strategy.

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.

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 (0.01 – 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 conduced 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.

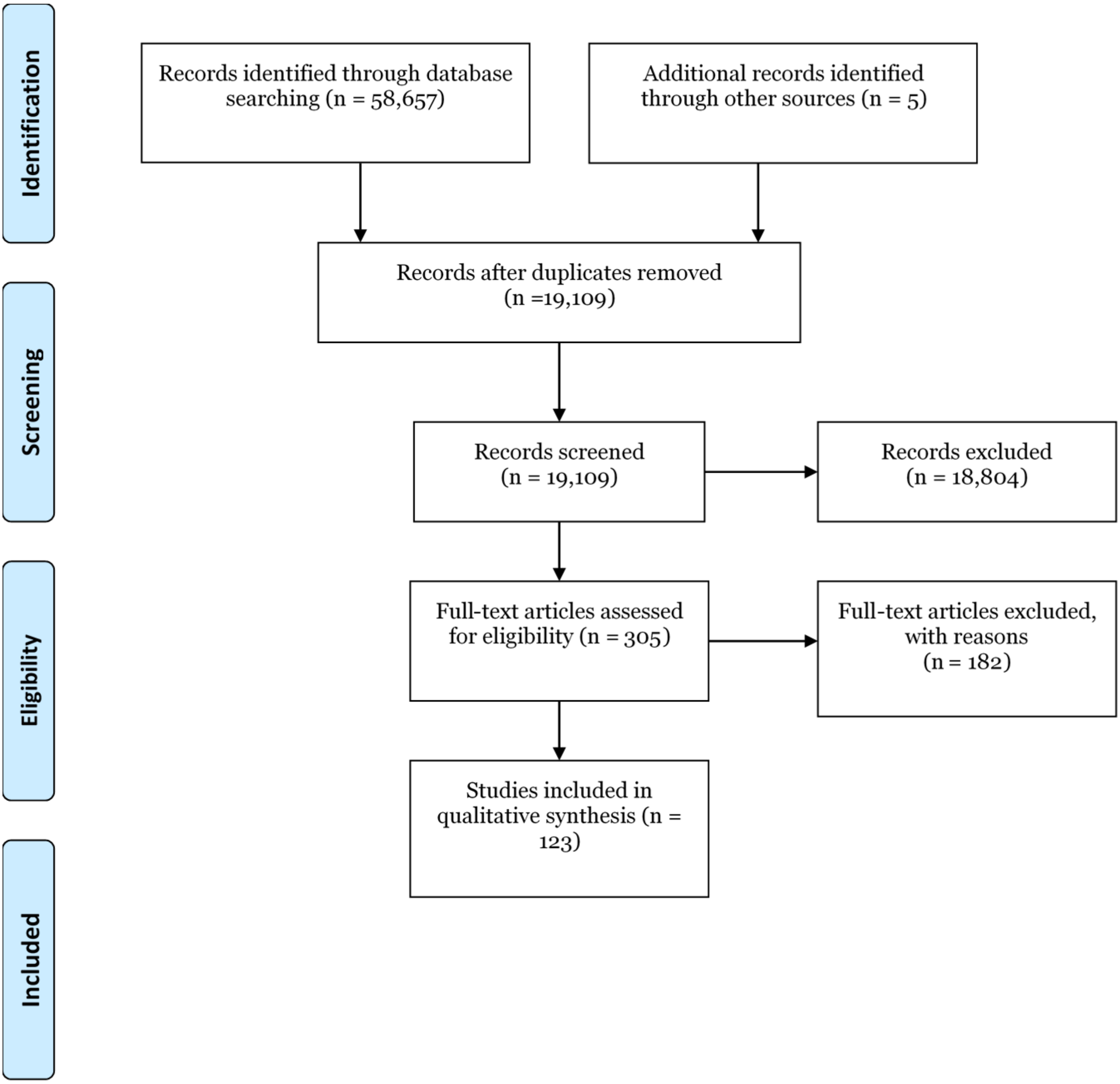

Figure 2. PRISMA diagram of article flow for the synthesis.

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.

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

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.

Reason for exclusion of impairment assessment tool	*Excluded impairment assessment tools*
Unlikely to influence riding performance/could be compensated by adaptations	*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*
Categorization of the severity of a health condition or were only suitable for one condition	*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*
Most items performed in standing	*Berg Balance Scale*
Already included in classification assessment	*Manual Muscle Test, range of movement*
Impractical to test in Para athletes	*Biering-Sørensen Test*
Construct already captured on other tools	*Scale for Contraversive Pushing and Modified Scale for Contraversive Pushing*
No studies of reliability or validity found	*Burke Lateropulsion Scale, Nottingham Assessment of Somato-sensations, Grasp and Release Test, Functional Axial Rotation, Catherine Bergogo Scale*

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.

Impairment Assessment Tool	*Population (sample size, age, health condition)*	*Usability Score*
Ashworth Scale *(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 & Ada, 1999)*	*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).*	*2*
Modified Ashworth Scale *(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 & Smith, 1987; Cooper et al., 2005; Craven & 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)*	*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).*	*2*
Re-Modified Ashworth Scale *(Ansari et al., 2012; Ansari et al., 2008b, 2009a, b; Ghotbi et al., 2009, 2011; Kaya et al., 2011; Mishra & Ganesh, 2014; Naghdi et al., 2007, 2008)*	*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).*	*2*
FIST *(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)*	*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), Hemicolectomy (n = 1), medically complex (n = 1), cardiac condition (n = 1).*	*1*
SARA *(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)*	*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)*	*3*

Impairment Assessment Tool	*Population (sample size, age, health condition)*	*Usability Score*
Tardieu Scale *(Paulis et al., 2011)*	*Overall sample size: n = 13* *Age (Mean, SD): 70.2, 12.3 years* *Condition: stroke (n = 13).*	*5*
Modified Tardieu Scale *(Akpinar et al., 2017; Ansari et al., 2008c, 2013; Azarnia et al. 2021; Ben-Shabat et al., 2013; Li et al., 2014; Mehrholz et al., 2005a; Naghdi et al., 2014, 2017; Santos et al. 2021; Singh et al., 2011; Sonvane & Kumar, 2019)*	*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 (Cabanas-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)	*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), fascio-scapulohumeral muscular dystrophy (n = 3), Becker muscular dystrophy (n = 3).*	*1*
Modified Trunk Impairment Scale *(YunBok et al., 2018)*	*Overall sample size: n = 55* *Age (Mean, SD, range) = 60.0, 2.5, 57 - 65 years* *Condition: stroke (n = 55).*	*1*
Hand-held Dynamometry *(Aguiar et al., 2016; Akshintala et al., 2021; Aksu & Yakut, 2003; Baschung Pfister et al., 2018; Beck et al., 1999; Bohannon, 1986, 1992, 1993, 1995; Bohannon et al., 2013; Bohannon & Andrews, 1987; Brinkmann, 1994; Busse et al., 2008; Cardin & Bohannon, 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)*	*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*

Impairment Assessment Tool	*Population (sample size, age, health condition)*	*Usability Score*
	muscular dystrophy (n = 1), myotonic dystrophy (n = 1), limb girdle muscular dystrophy (n = 2), fascia-scapulohumeral muscular dystrophy (n = 2), myositis (n = 4).	
Motricity Index *(Bohannon, 1995, 1999; Cameron & Bohannon, 2000; Collin & Wade, 1990; Fayazi et al., 2012; Vos-Vromans et al., 2005)*	*Overall sample size: n = 110* *Age (Mean, range) = 60.3, 15 - 81 years* *Conditions: stroke (n = 91), hemiparesis (n = 19).*	*3*
Trunk Control Test *(Collin & Wade, 1990; Fil Balkan et al., 2019; Franchignoni et al., 1997; Parlak Demir & Yildirim, 2015)*	*Overall sample size: n = 222* *Age (Mean, range) = 55.2, 15 - 83 years* *Conditions: stroke (n = 156), myopathy (n = 31), myotonic dystrophy (n = 22), limb girdle muscular dystrophy (n = 7), fascio-scapulohumeral muscular dystrophy (n = 3), Becker muscular dystrophy (n = 3).*	*2*

[1] Key: SARA- Scale for assessment and rating of ataxia; FIST- Function in Sitting Test. Overall usability score range = 0 (indicating excellent utility) – 11 (indicating low utility for Para dressage classification).

Table 4. Reliability and validity of included impairment assessment tools.

Impairment Assessment Tool	*Demographics*	*Reliability*	*Validity*	*Overall rating**
Ashworth Scale *(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 & Ada, 1999)*	*Total number of studies: 8,* *Total number of participants: 136* *Average age: 54.2*	*Interrater: ICC = 0.58, Kendall's Coefficient of Concordance = 0.92, Kendall's Tau = 0.39, Cohen's Weighted Kappa = 0.2* *Intrarater: Coefficient of Repeatability = 8, Coefficient of Variation = 6.4, Cohen's Weighted Kappa = 0.71*	*SRCC: EMG = 0.39*	*Reliability: Slight to substantial/moderate.* *Validity: Fair*
Modified Ashworth Scale *(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 & Smith, 1987; Cooper et al., 2005; Craven & 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*	*Total number of studies: 22,* *Total number of participants: 747* *Average age: 53.3*	*Interrater: ICC = 0.56, SRCC = 0.63, Kendall's Tau = 0.63, Cohen's Weighted Kappa = 0.32* *Intrarater: ICC = 0.61, SRCC = 0.65, Kendall's Tau = 0.06, Cohen's Weighted Kappa = 0.59* *Test-retest: ICC = 0.66* *Inter-session: ICC = 0.27* *Consistency: SRCC = 0.15, Cramer's V Nominal = 0.27*	*SRCC: EMG magnitude = 0.21*	*Reliability: Fair to moderate* *Validity: Poor*

Impairment Assessment Tool	*Demographics*	*Reliability*	*Validity*	*Overall rating**
al., 1992; Tederko et al., 2007; Zurawski et al., 2019)				
***Re-Modified Ashworth Scale** (Ansari et al., 2012; Ansari et al., 2008b, 2009a, b; Ghotbi et al., 2009, 2011; Kaya et al., 2011; Mishra & Ganesh, 2014; Naghdi et al., 2007, 2008)*	*Total number of studies: 10,* *Total number of participants: 270* *Average age: 50.6*	*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*	*SRCC: H-indexes = 0.58*	*Reliability: Almost perfect* *Validity: Moderately strong*
***FIST** (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)*	*Total number of studies: 8,* *Total number of participants: 376* *Average age: 56.6*	*Interrater: ICC = 0.98* *Intrarater: ICC = 0.98* *Test-retest: ICC = 0.95* *Consistency: Cronbach's Alpha = 0.93*	*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*	*Reliability: Excellent* *Validity: Poor to very strong*
***SARA** (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)*	*Total number of studies: 12,* *Total number of participants: 870* *Average age: 44.4*	*Interrater: ICC = 0.98, PCC = 0.86* *Intrarater: ICC = 0.97* *Test-retest: ICC = 0.98* *Consistency: Cronbach's Alpha = 0.90*	*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*	*Reliability: Excellent* *Validity: Moderate to very strong*
***Tardieu Scale** (Paulis et al., 2011)*	*Total number of studies: 1,* *Total number of participants: 13* *Average age: 70.2*	*Interrater: ICC = 0.75* *Test-retest: ICC = 0.84*		*Reliability: Good* *Validity: NA*

Impairment Assessment Tool	Demographics	Reliability	Validity	Overall rating*
Modified Tardieu Scale *(Akpinar et al., 2017; Ansari et al., 2008c, 2013; Azarnia et al. 2021; Ben-Shabat et al., 2013; Li et al., 2014; Mehrholz et al., 2005a; Naghdi et al., 2014, 2017; Singh et al., 2011; Santos et al., 2021; Sonvane & Kumar, 2019)*	*Total number of studies: 12, Total number of participants: 537 Average age: 53.7*	*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*	*SRCC: Hmax/Mmax = -0.04, Hslp/Mslp = 0.24, MAS = 0.81*	*Reliability: Moderate to substantial/good Validity: Poor to very strong*
Trunk Impairment Scale *(Cabanas-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)*	*Total number of studies: 14, Total number of participants: 1045 Average age: 60.4*	*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*	*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*	*Reliability: Excellent/almost perfect Validity: Moderately strong to very strong*
Modified Trunk Impairment Scale *(YunBok et al., 2018)*	*Total number of studies: 1, Total number of participants: 55 Average age: 60.0*		*SRCC: BBS = 0.82, Postural Assessment Scale = 0.55, TCT = 0.63*	*Reliability: NA Validity: Moderately strong to very strong*
Hand-held Dynamometry *(Aguiar et al., 2016; Akshintala et al., 2021; Aksu & Yakut, 2003; Baschung Pfister et al., 2018; Beck et al., 1999; Bohannon, 1986, 1992, 1993, 1995; Bohannon et al., 2013; Bohannon & Andrews, 1987; Brinkmann, 1994; Busse et al., 2008; Cardin & Bohannon, 2017; Dyball et al., 2011; Eken et al., 2020; Ekstrand et al., 2015; Faria et al., 2013; Goonetilleke et al., 1994;*	*Total number of studies: 35, Total number of participants: 1203 Average age: 50.7*	*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*	*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*	*Reliability: Good to excellent/very strong Validity: Poor to very strong*

Impairment Assessment Tool	Demographics	Reliability	Validity	Overall rating*
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)		Inter-session: ICC = 0.94, PCC = 0.96, SEM = 0.10 Parallel Forms Reliability: PCC = 0.82 Consistency: Cronbach's Alpha = 0.95		
Motricity Index (Bohannon, 1995, 1999; Cameron & Bohannon, 2000; Collin & Wade, 1990; Fayazi et al., 2012; Vos-Vromans et al., 2005)	Total number of studies: 6, Total number of participants: 110 Average age: 60.3	Interrater: SRCC = 0.88 Test-retest: SEM = 4.7	SRCC: RMA = 0.77 PCC: Dynamometry Value = 0.78	Reliability: Very strong. Validity: Moderately strong
Trunk Control Test (Collin & Wade, 1990; Fil Balkan et al., 2019; Franchignoni et al., 1997; Parlak Demir & Yildirim, 2015)	Total number of studies: 4, Total number of participants: 222 Average age: 55.2	Interrater: SRCC = 0.76 Intrarater: ICC = 0.98 Test-retest: SRCC = 0.96 Consistency: Cronbach's Alpha = 0.81	SRCC: RMA = 0.75, FIM = 0.50, TIS = 0.68	Reliability: Moderately strong to excellent Validity: Fair to moderately strong

[1] 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).

Discussion

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), Re-modified 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 stoke 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 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 out-perform 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; LBStgeorge@uclan.ac.uk, SJHobbs1@uclan.ac.uk, JAlexander3@uclan.ac.uk

* Correspondence: RStockley1@uclan.ac.uk; Tel.: +44-1772-894998

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

Abou, L., Sung, J., Sosnoff, J. J. & Rice, L. A. (2020). Reliability and validity of the function in sitting test among non-ambulatory individuals with spinal cord injury. *The Journal of Spinal Cord Medicine, 43*(6), 846-853. https://doi.org/10.1080/10790268.2019.1605749

Aguiar, L. T., Martins, J. C., Lara, E. M., Albuquerque, J. A., Teixeira-Salmela, L. F., & Faria, C. D. (2016). Dynamometry for the measurement of grip, pinch, and trunk muscles strength in subjects with subacute stroke: Reliability and different number of trials. *Brazilian Journal of Physical Therapy, 20*(5), 395–404. https://dx.doi.org/10.1590/bjpt-rbf.2014.0173

Akpinar, P., Atici, A., Ozkan, F. U., Aktas, I., Kulcu, D. G., Sari, A., & Durmus, B. (2017). Reliability of the Modified Ashworth Scale and Modified Tardieu Scale in patients with spinal cord injuries. *Spinal Cord, 55*(10), 944–949. https://dx.doi.org/10.1038/sc.2017.48

Akshintala, S., Khalil, N., Yohay, K., Muzikansky, A., Allen, J., Yaffe, A., Gross, A. M., Fisher, M. J., Blakeley, J. O., Oberlander, B. & Pudel, M. (2021). Reliability of Handheld Dynamometry to Measure Focal Muscle Weakness in Neurofibromatosis Types 1 and 2. *Neurology, 97*(7 Supplement 1), S99-S110. https://doi.org/10.1212/WNL.0000000000012439

Aksu, S., & Yakut, Y. (2003). Test-retest reliability of hand-held dynamometer and manual muscle strength measurements in amyotrophic lateral sclerosis. *Fizyoterapi Rehabilitasyon, 14*(2), 66–71. https://app.trdizin.gov.tr/publication/paper/detail/TXprMU5EazU

Allison, S. C., Abraham, L. D., & Petersen, C. L. (1996). Reliability of the Modified Ashworth Scale in the assessment of plantarflexor muscle spasticity in patients with traumatic brain injury. *International Journal of Rehabilitation Research, 19*(1), 67–78. https://doi.org/10.1097/00004356-199603000-00007

Altmann, V. C., Groen, B. E., Hart, A. L., Vanlandewijck, Y. C., & Keijsers, N. L. W. (2018). Classifying trunk strength impairment according to the activity limitation caused in wheelchair rugby performance. *Scandinavian Journal of Medicine & Science in Sports, 28*(2), 649–657. https://doi.org/10.1111/sms.12921

Altmann, V. C., Groen, B. E., Groenen, K. H., Vanlandewijck, Y. C., van Limbeek, J., & Keijsers, N. L. (2016). Construct validity of the trunk impairment classification system in relation to objective measures of trunk impairment. *Archives of Physical Medicine and Rehabilitation, 97*(3), 437-444. https://doi.org/10.1016/j.apmr.2015.10.096

Ansari, N. N., Naghdi, S., Hasson, S., Azarsa, M. H., & Azarnia, S. (2008c). The Modified Tardieu Scale for the measurement of elbow flexor spasticity in adult patients with hemiplegia. *Brain Injury, 22*(13–14), 1007–1012. https://dx.doi.org/10.1080/02699050802530557

Ansari, N. N., Naghdi, S., Hasson, S., Fakhari, Z., Mashayekhi, M., & Herasi, M. (2009a). Assessing the reliability of the Modified Ashworth Scale between two physiotherapists in adult patients with hemiplegia. *Neurorehabilitation, 25*(4), 235–240. https://dx.doi.org/10.3233/NRE-2009-0520

Ansari, N. N., Naghdi, S., Hasson, S., Mousakhani, A., Nouriyan, A., & Omidvar, Z. (2009b). Inter-rater reliability of the Modified Ashworth Scale as a clinical tool in measurements of post-stroke elbow flexor spasticity. *Neurorehabilitation, 24*(3), 225–229. https://dx.doi.org/10.3233/NRE-2009-0472

Ansari, N. N., Naghdi, S., Hasson, S., Rastgoo, M., Amini, M., & Forogh, B. (2013). Clinical assessment of ankle plantarflexor spasticity in adult patients after stroke: Inter-and intra-rater reliability of the Modified Tardieu Scale. *Brain Injury, 27*(5), 605–612. https://doi.org/10.3109/02699052.2012.750744

Ansari, N. N., Naghdi, S., Khosravian Arab, T., & Jalaie, S. (2008a). The interrater and intrarater reliability of the Modified Ashworth Scale in the assessment of muscle spasticity: Limb and muscle group effect. *Neurorehabilitation, 23*(3), 231–237. http://dx.doi.org/10.3233/NRE-2008-23304

Ansari, N. N., Naghdi, S., Mashayekhi, M., Hasson, S., Fakhari, Z., & Jalaie, S. (2012). Intra-rater reliability of the Modified Ashworth Scale (MMAS) in the assessment of upper-limb muscle spasticity. *Neurorehabilitation, 31*(2), 215–222. http://dx.doi.org/10.3233/NRE-2012-0791

Ansari, N. N., Naghdi, S., Moammeri, H., & Jalaie, S. (2006). Ashworth Scales are unreliable for the assessment of muscle spasticity. *Physiotherapy Theory and Practice, 22*(3), 119–125. http://dx.doi.org/10.1080/09593980600724188

Ansari, N. N., Naghdi, S., Younesian, P., & Shayeghan, M. (2008b). Inter- and intrarater reliability of the Modified Ashworth Scale in patients with knee extensor poststroke spasticity. *Physiotherapy Theory & Practice, 24*(3), 205–213. https://dx.doi.org/10.1080/09593980701523802

Azarnia, S., Abdollahi, I., Minoo, M., Naghdi, S. & Salavati, M. (2021). The inter-and intra-rater reliability of Modified Tardieu Scale in assessing spasticity of knee extensors in patients with multiple sclerosis. *Journal of Bodywork and Movement Therapies, 26*, 515-518. https://doi.org/10.1016/j.jbmt.2020.09.004

Bakers, J. N., van den Berg, L. H., Ajeks, T. G., Holleman, M. J., Verhoeven, J., Beelen, A., Visser-Meily, J., & van Eijk, R. (2021). Portable fixed dynamometry: towards remote muscle strength measurements in patients with motor neuron disease. *Journal of Neurology, 268*(5), 1738-1746. https://doi.org/10.1007/s00415-020-10366-9

Baschung Pfister, P., de Bruin, E. D., Sterkele, I., Maurer, B., de Bie, R. A. & Knols, R. H. (2018). Manual muscle testing and hand-held dynamometry in people with inflammatory myopathy: An intra-and interrater reliability and validity study. *PloS ONE, 13*(3), e0194531. https://doi.org/10.1371/journal.pone.0194531

Baunsgaard, C. B., Nissen, U. V., Christensen, K. B., & Biering-Sorensen, F. (2016). Modified Ashworth scale and spasm frequency score in spinal cord injury: Reliability and correlation. *Spinal Cord, 54*(9), 702–708. https://dx.doi.org/10.1038/sc.2015.230

Beck, M., Giess, R., Wurffel, W., Magnus, T., Ochs, G., & Toyka, K. V. (1999). Comparison of maximal voluntary isometric contraction and Drachman's hand-held dynamometry in evaluating patients with amyotrophic lateral sclerosis. *Muscle and Nerve, 22*(9), 1265–1270. http://dx.doi.org/10.1002/%28SICI%291097-4598%28199909%2922:9%3C1265::AID-MUS15%3E3.0.CO;2-F

Beckman, E. M., Connick, M. J., & Tweedy, S. M. (2017). Assessing muscle strength for the purpose of classification in Paralympic sport: A review and recommendations. *Journal of Science and Medicine in Sport, 20*(4), 391–396. https://doi.org/10.1016/j.jsams.2016.08.010

Ben-Shabat, E., Palit, M., Fini, N. A., Brooks, C. T., Winter, A., & Holland, A. E. (2013). Intra- and interrater reliability of the Modified Tardieu Scale for the assessment of lower limb spasticity in adults with neurologic injuries. *Archives of Physical Medicine and Rehabilitation, 94*(12), 2494–2501. https://doi.org/10.1016/j.apmr.2013.06.026

Blackburn, M., van Vliet, P., & Mockett, S. P. (2002). Reliability of measurements obtained with the modified Ashworth scale in the lower extremities of people with stroke. *Physical Therapy, 82*(1), 25–34. https://doi.org/10.1093/ptj/82.1.25

Bohannon, R. W. (1986). Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Physical Therapy, 66*(2), 206–209. https://doi.org/10.1093/ptj/66.2.206

Bohannon, R. W. (1992). Lateral trunk flexion strength: Impairment, measurement reliability and implications following unilateral brain lesion. *International Journal of Rehabilitation Research, 15*(3), 249–251. https://journals.lww.com/intjrehabilres/citation/1992/09000/lateral_trunk_flexion_strength___impairment,.10.aspx

Bohannon, R. W. (1993). Interday stability and validity of hand held dynamometer measurements obtained at bedside. *Journal of Physical Therapy Science, 5*(2), 61–66.

Bohannon, R. W. (1995). Consistency of paretic upper extremity motor performance soon after stroke. *Journal of Physical Therapy Science, 7*(2), 49–51. https://doi.org/10.1589/jpts.7.49

Bohannon, R. W. (1999). Motricity Index Scores are valid indicators of paretic upper extremity strength following stroke. *Journal of Physical Therapy Science, 11*(2), 59–61. https://doi.org/10.1589/jpts.11.59

Bohannon, R. W., & Andrews, A. W. (1987). Interrater reliability of hand-held dynamometry. *Physical Therapy, 67*(6), 931–933. https://doi.org/10.1093/ptj/67.6.931

Bohannon, R. W., Andrews, A. W., & Glenney, S. S. (2013). Responsiveness of measurements of lower-limb muscle strength obtained with a hand-held dynamometer from patients with stroke. *Isokinetics and Exercise Science, 21*(2), 129–134. http://dx.doi.org/10.3233/IES-130488

Bohannon, R. W., & Smith, M. B. (1987). Interrater reliability of a modified Ashworth scale of muscle spasticity. *Physical Therapy, 67*(2), 206–207. https://doi.org/10.1093/ptj/67.2.206

Bourcier, D., Bélanger, M., Côté, I., Brais, B., Synofzik, M., Brisson, J. D., Rodrigue, X., Gagnon, M. M., Mathieu, J. & Gagnon, C. (2020). Documenting the psychometric properties of the scale for the assessment and rating of ataxia to advance trial readiness of Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay. *Journal of the Neurological Sciences, 417*, 117050. https://doi.org/10.1016/j.jns.2020.117050

Braga-Neto, P., Godeiro-Junior, C., Dutra, L. A., Pedroso, J. L., & Barsottini, O. G. P. (2010). Translation and validation into Brazilian version of the Scale of the Assessment and Rating of Ataxia (SARA). *Arquivos de Neuro-Psiquiatria, 68*(2), 228–230. https://doi.org/10.1590/S0004-282X2010000200014

Brandsma, R., Lawerman, T. F., Kuiper, M. J., Lunsing, R. J., Burger, H., & Sival, D. A. (2017). Reliability and discriminant validity of ataxia rating scales in early onset ataxia. *Developmental Medicine & Child Neurology, 59*(4), 427–432. https://dx.doi.org/10.1111/dmcn.13291

Brashear, A., Zafonte, R., Corcoran, M., Galvez-Jimenez, N., Gracies, J., Gordon, M. F., McAfee, A., Ruffing, K., Thompson, B., Williams, M., Lee, C., & Turkel, C. (2002). Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity. *Archives of Physical Medicine & Rehabilitation, 83*(10), 1349–1354. https://doi.org/10.1053/apmr.2002.35474

Brinkmann, J. R. (1994). Comparison of a hand-held and fixed dynamometer in measuring strength of patients with neuromuscular disease. *Journal of Orthopaedic & Sports Physical Therapy, 19*(2), 100–104. https://doi.org/10.2519/jospt.1994.19.2.100

Bürk, K., Mälzig, U., Wolf, S., Heck, S., Dimitriadis, K., Schmitz-Hübsch, T., Hering, S., Lindig, T. M., Haug, V., Timmann, D., Degen, I., Kruse, B., Dörr, J. M., Ratzka, S., Ivo, A., Schöls, L., Boesch, S., Klockgether, T., Klopstock, T., & Schulz, J. B. (2009). Comparison of three clinical rating scales in Friedreich ataxia (FRDA). *Movement Disorders, 24*(12), 1779–1784. https://doi.org/10.1002/mds.22660

Burkett, B., Payton, C., Van de Vliet, P., Jarvis, H., Daly, D., Mehrkuehler, C., Kilian, M., & Hogarth, L. (2018). Performance characteristics of para swimmers. *Physical Medicine and Rehabilitation Clinics of North America, 29*(2), 333–346. https://doi.org/10.1016/j.pmr.2018.01.011

Busse, M. E., Hughes, G., Wiles, C. M., & Rosser, A. E. (2008). Use of hand-held dynamometry in the evaluation of lower limb muscle strength in people with Huntington's disease. *Journal of Neurology, 255*(10), 1534–1540. https://dx.doi.org/10.1007/s00415-008-0964-x

Byström, A., Roepstroff, L., Geser-von Peinen, K., Weishaupt, M.A. & Rhodin, M. (2015). Differences in rider movement pattern between different degrees of collection at the trot in high-level dressage horses ridden on a treadmill. *Human Movement Science*, *41*, 1–8. https://doi.org/10.1016/j.humov.2015.01.016

Cabanas-Valdes, R., Urrutia, G., Bagur-Calafat, C., Caballero-Gomez, F. M., German-Romero, A., & Girabent-Farres, M. (2016). Validation of the Spanish version of the Trunk Impairment Scale Version 2.0 (TIS 2.0) to assess dynamic sitting balance and coordination in post-stroke adult patients. *Topics in Stroke Rehabilitation, 23*(4), 225–232. https://dx.doi.org/10.1080/10749357.2016.1151662

Cameron, D., & Bohannon, R. W. (2000). Criterion validity of lower extremity Motricity Index scores. *Clinical Rehabilitation, 14*(2), 208–211. https://doi.org/10.1191%2F026921500675786655

Cardin, G., & Bohannon, R. W. (2017). Hip extension strength measured using hand-held dynamometry in a rehabilitation setting. *Isokinetics & Exercise Science, 25*(2), 157–160. https://doi.org/10.3233/IES-182161

Chan, Y. H. (2003). Biostatistics 104: Correlational analysis. *Singapore Medical Journal, 44*(12), 614–619. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.493.8362&rep=rep1&type=pdf

Collin, C., & Wade, D. (1990). Assessing motor impairment after stroke: A pilot reliability study. *Journal of Neurology, Neurosurgery & Psychiatry, 53*(7), 576–579. http://dx.doi.org/10.1136/jnnp.53.7.576

Connick, M. J., Beckman, E., & Tweedy, S. M. (2018a). Evolution and development of best practice in Paralympic classification. In I. Brittain & A. Beacom (Eds.), *The Palgrave Handbook of Paralympic Studies* (pp. 389–416). Palgrave Macmillan, UK. https://doi.org/10.1057/978-1-137-47901-3_18

Connick, M. J., Beckman, E., Vanlandewijck, Y., Malone, L. A., Blomqvist, S., & Tweedy, S. M. (2018b). Cluster analysis of novel isometric strength measures produces a valid and evidence-based classification structure for wheelchair track racing. *British Journal of Sports Medicine, 52*(17), 1123–1129. https://doi.org/10.1136/bjsports-2017-097558

Cooper, A., Musa, I. M., van Deursen, R., & Wiles, C. M. (2005). Electromyography characterization of stretch responses in hemiparetic stroke patients and their relationship with the Modified Ashworth scale. *Clinical Rehabilitation, 19*(7), 760–766. https://doi.org/10.1191%2F0269215505cr888oa

Craven, B. C., & Morris, A. R. (2010). Modified Ashworth scale reliability for measurement of lower extremity spasticity among patients with SCI. *Spinal Cord, 48*(3), 207–213. https://dx.doi.org/10.1038/sc.2009.107

Dasoju, V., Kovela, R. K., Tedla, J. S., Sangadala, D. R., & Reddy, R. S. (2021). Psychometric properties of trunk impairment scale in children with spastic diplegia. *Scientific Reports*, *11*(1), 1–7. https://doi.org/10.1038/s41598-021-98104-7 de Jong, R., Vanreusel, B. & van Driel, R. (2010). Relationships between mainstream participation rates and elite sport success in disability sports. *European Journal of Adapted Physical Activity, 3*(1), 18–29. https://doi.org/10.5507/euj.2010.002

Dyball, K. M., Taylor, N. F., & Dodd, K. J. (2011). Retest reliability of measuring hip extensor muscle strength in different testing positions in young people with cerebral palsy. *BMC Pediatrics, 11*. https://doi.org/10.1186/1471-2431-11-42

Eckardt, F., & Witte, K. (2016). Kinematic analysis of the rider according to different skill levels in sitting trot and canter. *Journal of Equine Veterinary Science*, *39*, 51–57. https://doi.org/10.1016/j.jevs.2015.07.022

Ehrenbrusthoff, K., Ryan, C. G., Grüneberg, C. and Martin, D. J. (2018). A systematic review and meta-analysis of the reliability and validity of sensorimotor measurement instruments in people with chronic low back pain. *Musculoskeletal Science and Practice*, 35, 73-83. https://doi.org/10.1016/j.msksp.2018.02.007

Eken, M. M., Lamberts, R. P., Koschnick, S., Du Toit, J., Veerbeek, B. E. & Langerak, N. G. (2020). Lower extremity strength profile in ambulatory adults with cerebral palsy and spastic diplegia: Norm values and reliability for hand-held dynamometry. *PM&R, 12*(6), 573-580. https://doi.org/10.1002/pmrj.12257

Ekstrand, E., Lexell, J., & Brogardh, C. (2015). Isometric and isokinetic muscle strength in the upper extremity can be reliably measured in persons with chronic stroke. *Journal of Rehabilitation Medicine, 47*(8), 706–713. https://dx.doi.org/10.2340/165019770-1990

Erol, B. N., Aydoğan Arslan, S., Yaşar, E., Keskin, E. D. & Demirci, C.S. (2021). Reliability and validity of the Turkish version of Function in Sitting Test (FIST-T) in stroke. *Topics in Stroke Rehabilitation, 29*(1), 49–57. https://doi.org/10.1080/10749357.2020.1871284.

Faria, C. D. C., Aguiar, L. T., Lara, E. M., Souza, L. A. Ce., Martins, J. C., & Teixeira-Salmela, L. F. (2013). Dynamometry for the assessment of grip, pinch, and trunk strength in subjects with chronic stroke: Reliability and various sources of outcome values. *International Journal of Physical Medicine & Rehabilitation, 1*(8), 1–5. https://doi.org/10.4172/2329-9096.1000168

Fayazi, M., Dehkordi, S. N., Dadgoo, M., & Salehi, M. (2012). Test-retest reliability of Motricity Index strength assessments for lower extremity in post stroke hemiparesis. *Medical Journal of the Islamic Republic of Iran, 26*(1), 27–30. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3587895/

Fédération Equestre Internationale (2022). *Para equestrian manual for classifiers*. Fédération Equestre Internationale. https://inside.fei.org/sites/default/files/FEI%20Manual%20for%20Classifiers%202020_update01.01.2022_0.pdf

Fédération Equestre Internationale (2017). *Para-equestrian classification manual.* Fédération Equestre Internationale. https://www.rda.org.uk/assets/FEI-Para-Equestrian-Classification-Manual-April-2017.pdf

Fil Balkan, A., Salcı, Y., Keklicek, H., Çetin, B., Adın, R. M. & Armutlu, K. (2019). The trunk control: Which scale is the best in very acute stroke patients? *Topics in Stroke Rehabilitation, 26*(5), 359-365. https://doi.org/10.1080/10749357.2019.1607994

Fleuren, J. F., Voerman, G. E., Erren-Wolters, C. V., Snoek, G. J., Rietman, J. S., Hermens, H. J., & Nene, A. V. (2010). Stop using the Ashworth Scale for the assessment of spasticity. *Journal of Neurology, Neurosurgery & Psychiatry, 81*(1), 46–52. https://dx.doi.org/10.1136/jnnp.2009.177071

Franchignoni, F. P., Tesio, L., Ricupero, C., & Martino, M. T. (1997). Trunk control test as an early predictor of stroke rehabilitation outcome. *Stroke, 28*(7), 1382–1385. https://doi.org/10.1161/01.STR.28.7.1382

Ganguly, J., Kulshreshtha, D., Almotiri, M., & Jog, M. (2021). Muscle tone physiology and abnormalities. *Toxins*, *13*(4), 282. https://doi.org/10.3390%2Ftoxins13040282

Garg, N., Park, S. B., Vucic, S., Yiannikas, C., Spies, J., Howells, J., Huynh, W., Matamala, J. M., Krishnan, A. V., Pollard, J. D., Cornblath, D. R. Reilly, M. M. & Kiernan, M. C. (2017). Differentiating lower motor neuron syndromes. *Journal of Neurology,* *Neurosurgery & Psychiatry*, *88*(6), 474-483. http://dx.doi.org/10.1136/jnnp-2016-313526

Ghotbi, N., Ansari, N. N., Naghdi, S., & Hasson, S. (2011). Measurement of lower-limb muscle spasticity: Intrarater reliability of Modified Ashworth Scale. *Journal of Rehabilitation Research & Development, 48*(1), 83–88. https://doi.org/10.1682/JRRD.2010.02.0020

Ghotbi, N., Ansari, N. N., Naghdi, S., Hasson, S., Jamshidpour, B., & Amiri, S. (2009). Inter-rater reliability of the Modified Ashworth Scale in assessing lower limb muscle spasticity. *Brain Injury, 23*(10), 815–819. https://dx.doi.org/10.1080/02699050903200548

Goonetilleke, A., Modarres-Sadeghi, H., & Guiloff, R. J. (1994). Accuracy, reproducibility, and variability of hand-held dynamometry in motor neuron disease. *Journal of Neurology, Neurosurgery & Psychiatry, 57*(3), 326–332. http://dx.doi.org/10.1136/jnnp.57.3.326

Gorman, S. L., Harro, C. C., Platko, C., & Greenwald, C. (2014a). Examining the function in sitting test for validity, responsiveness, and minimal clinically important difference in inpatient rehabilitation. *Archives of Physical Medicine & Rehabilitation, 95*(12), 2304–2311. https://dx.doi.org/10.1016/j.apmr.2014.07.415

Gorman, S. L., Radtka, S., Melnick, M. E., Abrams, G. M., & Byl, N. N. (2010). Development and validation of the Function in Sitting Test in adults with acute stroke. *Journal of Neurologic Physical Therapy, 34*(3), 150–160. https://doi.org/10.1097/npt.0b013e3181f0065f

Gorman, S. L., Rivera, M., & McCarthy, L. (2014b). Reliability of the Function in Sitting Test (FIST). *Rehabilitation Research & Practice, 2014*, 593280. https://dx.doi.org/10.1155/2014/593280

Gregson, J. M., Leathley, M. J., Moore, A. P., Sharma, A. K., Smith, T. L., & Watkins, C. L. (1999). Reliability of the Tone Assessment Scale and the modified Ashworth scale as clinical tools for assessing poststroke spasticity. *Archives of Physical Medicine & Rehabilitation, 80*(9), 1013–1016. https://doi.org/10.1016/S0003-9993(99)90053-9

Gregson, J. M., Leathley, M. J., Moore, A. P., Smith, T. L., Sharma, A. K., & Watkins, C. L. (2000). Reliability of measurements of muscle tone and muscle power in stroke patients. *Age & Ageing, 29*(3), 223–228. https://doi.org/10.1093/ageing/29.3.223

Haas, B. M., Bergstrom, E., Jamous, A., & Bennie, A. (1996). The inter rater reliability of the original and of the modified Ashworth scale for the assessment of spasticity in patients with spinal cord injury. *Spinal Cord, 34*(9), 560–564. https://doi.org/10.1038/sc.1996.100

Harb, A., & Kishner, S. (2022). Modified Ashworth Scale. In *StatPearls [Internet]*. Florida: StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK554572/

Hayes, K., Walton, J. R., Szomor, Z. L., & Murrell, G. A. C. (2002). Reliability of 3 methods for assessing shoulder strength. *Journal of Shoulder and Elbow Surgery, 11*(1), 33–39. http://dx.doi.org/10.1067/mse.2002.119852

Higgins J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V.A., (Eds.). (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (2nd Ed). Chichester: John Wiley & Sons. www.training.cochrane.org/handbook

Hobbs, S. J., St George, L., Reed, J., Stockley, R., Thetford, C., Sinclair, J., Williams, J., Nankervis, K., & Clayton, H. M. (2020). A scoping review of determinants of performance in dressage. *PeerJ, 8,* e9022. https://doi.org/10.7717/peerj.9022

Hogarth, L., Nicholson, V., Spathis, J., Tweedy, S., Beckman, E., Connick, M., van de Vliet, P., Payton, C., & Burkett, B. (2019). A battery of strength tests for evidence-based classification in Para swimming. *Journal of Sports Sciences, 37*(4), 404–413. https://doi.org/10.1080/02640414.2018.1504606

International Paralympic Committee (2015). *2015 Athlete Classification Code*. International Paralympic Committee. https://www.paralympic.org/classification-code

Kaya, T., Goksel Karatepe, A., Gunaydin, R., Koc, A., & Altundal Ercan, U. (2011). Inter-rater reliability of the Modified Ashworth Scale and modified Modified Ashworth Scale in assessing poststroke elbow flexor spasticity. *International Journal of Rehabilitation Research, 34*(1), 59–64. https://doi.org/10.1097/MRR.0b013e32833d6cdf

Kilmer, D. D., McCrory, M. A., Wright, N. C., Rosko, R., Kim, H., & Aitkens, S. G. (1997). Hand held dynamometery reliability in persons with neuropathic weakness. *Archives of Physical Medicine & Rehabilitation, 78*(12), 1364–1368. https://doi.org/10.1016/S0003-9993(97)90311-7

Kim, B.-R., Lee, J.-Y., Kim, M. J., Jung, H., & Lee, J. (2014). Korean version of the scale for the assessment and rating of ataxia in ataxic stroke patients. *Annals of Rehabilitation Medicine, 38*(6), 742–751. https://doi.org/10.5535%2Farm.2014.38.6.742

Knak, K.L., Sheikh, A.M., Andersen, H., Witting, N. & Vissing, J. (2020). Intrarater reliability and validity of outcome measures in myotonic dystrophy type 1. *Neurology, 94*(24), e2508-e2520. https://doi.org/10.1212/WNL.0000000000009625

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Larson, C. A., Tezak, W. D., Malley, M. S., & Thornton, W. (2010). Assessment of postural muscle strength in sitting: Reliability of measures obtained with hand-held dynamometry in individuals with spinal cord injury. *Journal of Neurologic Physical Therapy, 34*(1), 24–31. https://doi.org/10.1097/NPT.0b013e3181cf5c49

Lee, K. C., Carson, L., Kinnin, E., & Patterson, V. (1989). The Ashworth Scale: A reliable and reproducible method of measuring spasticity. *Journal of Neurologic Rehabilitation, 3*(4), 205–209. https://doi.org/10.1177%2F136140968900300406

Li, F., Wu, Y., & Li, X. (2014). Test-retest reliability and inter-rater reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in hemiplegic patients with stroke. *European Journal of Physical and Rehabilitation Medicine, 50*(1), 9–15. https://pubmed.ncbi.nlm.nih.gov/24309501/

Livesley, E. (1992). The intra-observer reliability of the hand-held myometer in the measurement of isotonic muscle strength in chronic spasticity. *Physiotherapy, 78*(12), 918–921. https://doi.org/10.1016/S0031-9406(10)60502-0

Lombardi, B., Orioli, A., Casavola, D., & Paci, M. (2017). The Italian version of the Trunk Impairment Scale: Development and psychometric properties. *European Journal of Physical and Rehabilitation Medicine, 53*(4), 516–520. https://dx.doi.org/10.23736/S1973-9087.17.04371-4

Mann, D. L., Tweedy, S. M., Jackson, R. C., & Vanlandewijck, Y. C. (2021). Classifying the evidence for evidence-based classification in Paralympic sport. *Journal of Sports Sciences, 39*(s1), 1–6. https://doi.org/10.1080/02640414.2021.1955523

May, L. A., Burnham, R. S., & Steadward, R. D. (1997). Assessment of isokinetic and hand-held dynamometer measures of shoulder rotator strength among individuals with spinal cord injury. *Archives of Physical Medicine & Rehabilitation, 78*(3), 251–255. https://doi.org/10.1016/S0003-9993(97)90029-0

Mehrholz, J., Major, Y., Meiçner, D., Sandi-Gahun, S., Koch, R., & Pohl, M. (2005b). The influence of contractures and variation in measurement stretching velocity on the reliability of the Modified Ashworth Scale in patients with severe brain injury. *Clinical Rehabilitation, 19*(1), 63–72. https://doi.org/10.1191%2F0269215505cr824oa

Mehrholz, J., Wagner, K., Mei, ßner, D., Grundmann, K., Zange, C., Koch, R., & Pohl, M. (2005a). Reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in adult patients with severe brain injury: a comparison study. *Clinical Rehabilitation, 19*(7), 751–759. https://doi.org/10.1191%2F0269215505cr889oa

Mentiplay, B. F., Tan, D., Williams, G., Adair, B., Pua, Y. H., Bower, K. J., & Clark, R. A. (2018). Assessment of isometric muscle strength and rate of torque development with hand-held dynamometry: Test-retest reliability and relationship with gait velocity after stroke. *Journal of Biomechanics, 75*, 171–175. https://dx.doi.org/10.1016/j.jbiomech.2018.04.032

Meseguer-Henarejos, A. B., Sanchez-Meca, J., Lopez-Pina, J. A., & Carles-Hernandez, R. (2017). Inter-and intra-rater reliability of the Modified Ashworth Scale: a systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, *54*(4), 576–590. https://doi.org/10.23736/s1973-9087.17.04796-7

Mishra, C., & Ganesh, G. S. (2014). Inter-rater reliability of modified Modified Ashworth Scale in the assessment of plantar flexor muscle spasticity in patients with spinal cord injury. *Physiotherapy Research International, 19*(4), 231–237. https://dx.doi.org/10.1002/pri.1588

Monticone, M., Ambrosini, E., Verheyden, G., Brivio, F., Brunati, R., Longoni, L., Mauri, G., Molteni, A., Nava, C., Rocca, B., & Ferrante, S. (2017). Development of the Italian version of the trunk impairment scale in subjects with acute and chronic stroke. Cross-cultural adaptation, reliability, validity and responsiveness. *Disability & Rehabilitation, 49*(1), 66–73. https://dx.doi.org/10.1080/09638288.2017.1373409

Moreno-Navarro, P., Manca, A., Martinez, G., Ventura, L., Barbado, D., Vera-García, F.J. & Deriu, F. (2021). Test-retest reliability and known-groups validity of trunk muscle tests in people with multiple sclerosis: A cross-sectional, case-control study. *Physical Therapy, 101*(5), 1-9. https://doi.org/10.1093/ptj/pzab049

Morris, S., Dodd, K., & Morris, M. (2008). Reliability of dynamometry to quantify isometric strength following traumatic brain injury. *Brain Injury, 22*(13–14), 1030–1037. http://dx.doi.org/10.1080/02699050802533775

Naghdi, S., Ansari, N. N., Abolhasani, H., Mansouri, K., Ghotbi, N., & Hasson, S. (2014). Electrophysiological evaluation of the Modified Tardieu Scale (MTS) in assessing poststroke wrist flexor spasticity. *Neurorehabilitation, 34*(1), 177–184. https://dx.doi.org/10.3233/NRE-131016

Naghdi, S., Ansari, N., Ghorbani-rad, S., Senobari, M., Sahraian, M., Ansari, N. N., & Sahraian, M. A. (2017). Intra-rater reliability of the Modified Tardieu Scale in patients with multiple sclerosis. *Neurological Sciences, 38*(1), 93–99. https://doi.org/10.1007/s10072-016-2714-7

Naghdi, S., Ansari, N. N., Azarnia, S., & Kazemnejad, A. (2008). Interrater reliability of the Modified Modified Ashworth Scale (MMAS) for patients with wrist flexor muscle spasticity. *Physiotherapy Theory & Practice, 24*(5), 372–379. https://dx.doi.org/10.1080/09593980802278959

Naghdi, S., Ebrahimi, I., Asgari, A., Olyaei, G. R., Kazemnejad, A., Mansouri, K., & Ansari, N. N. (2007). A preliminary study into the criterion validity of the modified Modified Ashworth Scale using the new measure of the alpha motoneuron excitability in spastic hemiplegia. *Electromyography & Clinical Neurophysiology, 47*(3), 187–192. https://pubmed.ncbi.nlm.nih.gov/17557652/

Nakhostin-Ansari, N., Naghdi, S., Moammeri, H., & Jalaie, S. (2006). A comparative study on the inter-rater reliability of the Ashworth Scales in assessment of spasticity. *Acta Medica Iranica, 44*(4), 246–250. https://acta.tums.ac.ir/index.php/acta/article/view/3192

Nicholson, V. P., Spathis, J. G., Hogarth, L. W., Connick, M. J., Beckman, E. M., Tweedy, S. M., Payton, C. J., & Burkett, B.J. (2018). Establishing the reliability of a novel battery of range of motion tests to enable evidence-based classification in Para Swimming. *Physical Therapy in Sport*, *32*, 34–41. https://doi.org/10.1016/j.ptsp.2018.04.021

Noreau, L., & Vachon, J. (1998). Comparison of three methods to assess muscular strength in individuals with spinal cord injury. *Spinal Cord, 36*, 716–723. https://doi.org/10.1038/sj.sc.3100646

Nuyens, G., De Weerdt, W., Ketalaer, P., Feys, H., De Wolf, L., Hantson, L., Nieuwboer, A., Spaepen, A., & Carton, H. (1994). Inter-rater reliability of the Ashworth scale in multiple sclerosis. *Clinical Rehabilitation, 8*(4), 286–292. https://doi.org/10.1177%2F026921559400800403

Olivier, A., Faugloire, E., Lejeune, L., Biau, S., & Isableu, B. (2017). Head stability and head-trunk coordination in horseback riders: the contribution of visual information according to expertise. *Frontiers in Human Neuroscience*, *11*, Article 11. https://doi.org/10.3389/fnhum.2017.00011

Paix, E. L., Tweedy, S. M., Connick, M. J., & Beckman, E. M. (2021). Differentiating maximal and submaximal voluntary strength measures for the purposes of medico-legal assessments and para sport classification: A systematic review. *European*

Journal of Sport Science, *21*(11), 1518–1550. https://doi.org/10.1080/17461391.2020.1834623

Palermo, A. E., Cahalin, L. P., Garcia, K. L. & Nash, M. S. (2020). Psychometric testing and clinical utility of a modified version of the function in sitting test for individuals with chronic spinal cord injury. *Archives of Physical Medicine and Rehabilitation, 101*(11), 1961-1972. https://doi.org/10.1016/j.apmr.2020.06.014

Pandyan, A. D., Price, C. I., Barnes, M. P., & Johnson, G. R. (2003). A biomechanical investigation into the validity of the modified Ashworth Scale as a measure of elbow spasticity. *Clinical Rehabilitation, 17*(3), 290–293. https://doi.org/10.1191%2F0269215503cr610oa

Pandyan, A. D., Price, C. I. M., Rodgers, H., Barnes, M. P., & Johnson, G. R. (2001). Biomechanical examination of a commonly used measure of spasticity. *Clinical Biomechanics, 16*(10), 859–865. http://dx.doi.org/10.1016/S0268-0033%2801%2900084-5

Parlak Demir, Y. P. & Yildirim, S. A. (2015). Reliability and validity of Trunk Control Test in patients with neuromuscular diseases. *Physiotherapy Theory & Practice, 31*(1), 39–44. https://dx.doi.org/10.3109/09593985.2014.945673

Parlak Demir, Y. P. & Yıldırım, S. A. (2018). Assessment of trunk control in patients with neuromuscular diseases: validity and reliability of the Trunk Impairment Scale. *Turkish Journal of Neurology*. *24*(2), 130-136. http://dx.doi.org/10.4274/tnd.36024

Paulis, W. D., Horemans, H. L. D., Brouwer, B. S., & Stam, H. J. (2011). Excellent test-retest and inter-rater reliability for Tardieu Scale measurements with inertial sensors in elbow flexors of stroke patients. *Gait and Posture, 33*(2), 185–189. http://dx.doi.org/10.1016/j.gaitpost.2010.10.094

Peham, C., Licka, T., Kapaun, M., & Scheidl, M. (2001). A new method to quantify harmony of the horse–rider system in dressage. *Sports Engineering*, *4*(2), 95–101. https://doi.org/10.1046/j.1460-2687.2001.00077.x

Quinzanos, J., Villa, A. R., Flores, A. A., & Perez, R. (2014). Proposal and validation of a clinical trunk control test in individuals with spinal cord injury. *Spinal Cord, 52*(6), 449–454. http://dx.doi.org/10.1038/sc.2014.34

Riddle, D. L., Finucane, S. D., Rothstein, J. M., & Walker, M. L. (1989). Intrasession and intersession reliability of hand-held dynamometer measurements taken on brain-damaged patients. *Physical Therapy, 69*(3), 182–189. https://doi.org/10.1093/ptj/69.3.182

Roldan, A., Barbado, D., Vera-Garcia, F. J., Sarabia, J. M., & Reina, R. (2020). Inter-rater reliability, concurrent validity and sensitivity of current methods to assess trunk function in boccia players with cerebral palsy. *Brain Sciences*, *10*(3), 130. https://doi.org/10.3390/brainsci10030130

Sag, S., Buyukavci, R., Sahin, F., Sag, M.S., Dogu, B. & Kuran, B. (2019). Assessing the validity and reliability of the Turkish version of the Trunk Impairment Scale in stroke patients. *Northern Clinics of Istanbul, 6*(2), 156-165. https://doi.org/10.14744%2Fnci.2018.01069

Salci, Y., Fil, A., Keklicek, H., Cetin, B., Armutlu, K., Dolgun, A., Tuncer, A., & Karabudak, R. (2017). Validity and reliability of the International Cooperative Ataxia Rating Scale (ICARS) and the Scale for the Assessment and Rating of Ataxia (SARA) in multiple sclerosis patients with ataxia. *Multiple Sclerosis and Related Disorders, 18*, 135–140. https://dx.doi.org/10.1016/j.msard.2017.09.032

Santos, P. L. D. A., Gaspar, R. C., Padula, N., Almeida, D. M. and Voos, M. C. (2021). Translation and cross-cultural adaptation to Brazilian Portuguese of the Modified Tardieu Scale for muscle tone assessment among patients with spinal cord injury. *Arquivos de Neuro-Psiquiatria, 79*(7), 590-597. https://doi.org/10.1590/0004-282X-anp-2020-0366

Saygin, D., Oddis, C.V., Moghadam-Kia, S., Rockette-Wagner, B., Neiman, N., Koontz, D. & Aggarwal, R. (2021). Hand-held dynamometry for assessment of muscle strength in patients with inflammatory myopathies. *Rheumatology, 60*(5), 2146-2156. https://doi.org/10.1093/rheumatology/keaa419

Schmitz-Hübsch, T., du Montcel, S. T., Baliko, L., Berciano, J., Boesch, S., Depondt, C., Giunti, P., Globas, C., Infante, J., Kang, J. S., Kremer, B., Mariotti, C., Melegh, B., Pandolfo, M., Rakowicz, M., Ribai, P., Rola, R., Schöls, L., Szymanski, S., & van de Warrenburg, B. P. (2006). Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology, 66*(11), 1717–1720. https://doi.org/10.1212/01.wnl.0000219042.60538.92
Schmitz-Hübsch, T., Fimmers, R., Rakowicz, M., Rola, R., Zdzienicka, E., Fancellu, R., Mariotti, C., Linnemann, C., Schöls, L., Timmann, D., Filla, A., Salvatore, E., Infante, J., Giunti, P., Labrum, R., Kremer, B., van de Warrenburg, B. P., Baliko, L., Melegh, B., & Depondt, C. (2010). Responsiveness of different rating instruments in spinocerebellar ataxia patients. *Neurology, 74*(8), 678–684. https://doi.org/10.1212/WNL.0b013e3181d1a6c9
Schwartz, S., Cohen, M. E., Herbison, G. J., & Shah, A. (1992). Relationship between two measures of upper extremity strength: Manual muscle test compared to hand-held myometry. *Archives of Physical Medicine & Rehabilitation, 73*(11), 1063–1068. https://doi.org/10.5555/uri:pii:000399939290172S
Seo, H. D., Kim, N. J., & Chung, Y. J. (2008). Reliability of the Korean version of the trunk impairment scale in patients with stroke. *Physical Therapy Korea, 15*(4), 87–96. https://www.koreascience.or.kr/article/JAKO200816549675769.view?orgId=anpor&hide=breadcrumb,journalinfo
Simpson, D., Ehrensberger, M., Nulty, C., Regan, J., Broderick, P., Blake, C., & Monaghan, K. (2019). Peak torque, rate of torque development and average torque of isometric ankle and elbow contractions show excellent test–retest reliability. *Hong Kong Physiotherapy Journal*, *39*(01), 67-76. https://doi.org/10.1142%2FS1013702519500069
Singh, P., Joshua, A., Ganeshan, S., & Suresh, S. (2011). Intra-rater reliability of the modified Tardieu scale to quantify spasticity in elbow flexors and ankle plantar flexors in adult stroke subjects. *Annals of Indian Academy of Neurology, 14*(1), 23–26. http://dx.doi.org/10.4103/0972-2327.78045
Sloan, R. L., Sinclair, E., Thompson, J., Taylor, S., & Pentland, B. (1992). Inter-rater reliability of the modified Ashworth Scale for spasticity in hemiplegic patients. *International Journal of Rehabilitation Research, 15*(2), 158–161. https://doi.org/10.1097/00004356-199206000-00009
Smith, R., Connick, M., Beckman, E., Hogarth, L., & Nicholson, V. (2021). Establishing the reliability of instrumented trunk impairment assessment methods to enable evidence-based classification in Para swimming. *Journal of Sports Sciences*, *39*(S1), 73-80. https://doi.org/10.1080/02640414.2021.1930699
Sonvane, I. & Kumar T. S. (2019). Test retest reliability of Modified Tardieu Scale to quantify the spasticity in elbow flexors in patients with cerebro vascular accident. *Indian Journal of Physiotherapy & Occupational Therapy, 13*(2), 38-42. https://www.doi.org/10.5958/0973-5674.2019.00042.X
St. George, L., Thetford, C., Clayton, H. M., & Hobbs, S. J. (2021). An exploration of stakeholder perceptions to inform the development of an evidence-based classification system in para dressage. *Journal of Sports Sciences, 40*(4), 459–469. https://doi.org/10.1080/02640414.2021.1997012
Streiner, D. L. (2015). *Health measurement scales: A practical guide to their development and use. 5th Ed.* Oxford University Press. https://doi.org/10.1093/med/9780199685219.001.0001
Sung, J., Ousley, C. M., Shen, S., Isaacs, Z. J., Sosnoff, J. J., & Rice, L. A. (2016). Reliability and validity of the function in sitting test in nonambulatory individuals with multiple sclerosis. *International Journal of Rehabilitation Research, 39*(4), 308–312. https://doi.org/10.1097/MRR.0000000000000188
Tan, S., Niu, H. X., Zhao, L., Gao, Y., Lu, J. M., Shi, C. H., Avinash, C., Wang, R. H., & Xu, Y. M. (2013). Reliability and validity of the Chinese version of the Scale for Assessment and Rating of Ataxia. *Chinese Medical Journal, 126*(11), 2045–2048. https://doi.org/10.3760/cma.j.issn.0366-6999.20121703

Tederko, P., Krasuski, M., Czech, J., Dargiel, A., Garwacka-Jodzis, I., & Wojciechowska, A. (2007). Reliability of clinical spasticity measurements in patients with cervical spinal cord injury. *Ortopedia Traumatologia Rehabilitacja, 9*(5), 467–483. https://pubmed.ncbi.nlm.nih.gov/18026067/

Tsai, Y. J., Tu, Y. K., Hsiao, C. K., & Su, F. C. (2015). Within-session reliability and smallest real difference of muscle strength following nerve transfers in patients with brachial plexus injuries. *Journal of Hand Surgery, 40*(6), 1196–1201. https://dx.doi.org/10.1016/j.jhsa.2015.01.043

Tweedy, S. M., Beckman, E. M., & Connick, M. J. (2014). Paralympic classification: Conceptual basis, current methods, and research update. *PM&R, 6*(8S), S11–S17. https://doi.org/10.1016/j.pmrj.2014.04.013

Tweedy, S. M., & Vanlandewijck, Y. C. (2011). International Paralympic Committee position stand—Background and scientific principles of classification in Paralympic sport. *British Journal of Sports Medicine, 45*(4), 259–269. https://doi.org/10.1136/bjsm.2009.065060

Tweedy, S. M., Williams, G., & Bourke, J. (2010). Selecting and modifying methods of manual muscle testing for classification in paralympic sport. *European Journal of Adapted Physical Activity, 3*(2), 7–16. https://doi.org/10.5507/euj.2010.005

Tweedy, S. M., Mann, D., & Vanlandewijck, Y. C. (2016). Research needs for the development of evidence-based systems of classification for physical, vision and intellectual impairments. In Y. C. Vanlandewijck & W. R. Thompson (Eds.), *Training and coaching the Paralympic Athlete* (pp. 122-149). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119045144.ch7

Tyson, S. F. & Connell, L. A. (2009). The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. *Clinical Rehabilitation, 23*(11), 1018–1033. https://doi.org/10.1177%2F0269215509339004

Tyson, S. F. & Connell, L. A. (2009). How to measure balance in clinical practice. A systematic review of the psychometrics and clinical utility of measures of balance activity for neurological conditions. *Clinical Rehabilitation, 23*(9), 824–840. https://doi.org/10.1177%2F0269215509335018

van de Pol, R. J., van Trijffel, E., & Lucas, C. (2010). Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. *Journal of Physiotherapy*, *56*(1), 7-17. https://doi.org/10.1016/S1836-9553(10)70049-7

van Langeveld, S. H., Aufdemkampe, G., & van Asbeck, F. W. A. (1996). Reliability of force measurement with a hand-held dynamometer in healthy subjects and force measurements in patients with poliomyelitis anterior acuta. *Journal of Rehabilitation Sciences, 9*(1), 2–10.

Vandervelde, L., Van den Bergh, P. Y. K., Renders, A., Goemans, N., & Thonnard, J.-L. (2009). Relationships between motor impairments and activity limitations in patients with neuromuscular disorders. *Journal of Neurology, Neurosurgery & Psychiatry, 80*(3), 326–332. https://doi.org/10.1136/jnnp.2008.150060

Vattanasilp, W., & Ada, L. (1999). The relationship between clinical and laboratory measures of spasticity. *Australian Journal of Physiotherapy, 45*(2), 135–139. https://doi.org/10.1016/S0004-9514(14)60345-7

Verheyden, G., Hughes, J., Jelsma, J., Nieuwboer, A., & De Weerdt, W. (2006a). Assessing motor impairment of the trunk in patients with traumatic brain injury: Reliability and validity of the Trunk Impairment Scale. *South African Journal of Physiotherapy, 62*(2), 23–28. https://doi.org/10.4102/sajp.v62i2.153

Verheyden, G., Nieuwboer, A., Feys, H., Thijs, V., Vaes, K., & De Weerdt, W. (2005). Discriminant ability of the Trunk Impairment Scale: A comparison between stroke patients and healthy individuals. *Disability and Rehabilitation, 27*(17), 1023–1028. https://doi.org/10.1080/09638280500052872

Verheyden, G., Nieuwboer, A., Martin, J., Preger, R., Kiekens, C., & De Weerdt, W. (2004). The Trunk Impairment Scale: A new tool to measure motor impairment of the trunk

after stroke. *Clinical Rehabilitation, 18*(3), 326–334. https://doi.org/10.1191%2F0269215504cr733oa
Verheyden, G., Nuyens, G., Nieuwboer, A., Van Asch, P., Ketelaer, P., & De Weerdt, W. (2006b). Reliability and validity of trunk assessment for people with multiple sclerosis. *Physical Therapy, 86*(1), 66–76. https://doi.org/10.1093/ptj/86.1.66
Verheyden, G., Willems, A. M., Ooms, L., & Nieuwboer, A. (2007). Validity of the trunk impairment scale as a measure of trunk performance in people with Parkinson's disease. *Archives of Physical Medicine & Rehabilitation, 88*(10), 1304–1308. https://doi.org/10.1016/j.apmr.2007.06.772
Visser, J., Mans, E., De Visser, M., Van Den Berg-Vos, R. M., Franssen, H., De Jong, J. M. B. V., Van Den Berg, L. H., Wokke, J. H. J., & De Haan, R. J. (2003). Comparison of maximal voluntary isometric contraction and hand-held dynamometry in measuring muscle strength of patients with progressive lower motor neuron syndrome. *Neuromuscular Disorders, 13*(9), 744–750. http://dx.doi.org/10.1016/S0960-8966%2803%2900135-4
Vos-Vromans, D. C. W. M., De Bie, R. A., Erdmann, P. G., & Van Meeteren, N. L. U. (2005). The responsiveness of the Ten-Meter Walking Test and other measures in patients with hemiparesis in the acute phase. *Physiotherapy Theory and Practice, 21*(3), 173–180. http://dx.doi.org/10.1080/09593980500212920
Weyer, A., Abele, M., Schmitz-Hubsch, T., Schoch, B., Frings, M., Timmann, D., & Klockgether, T. (2007). Reliability and validity of the scale for the assessment and rating of ataxia: A study in 64 ataxia patients. *Movement Disorders, 22*(11), 1633–1637. https://doi.org/10.1002/mds.21544
World Health Organisation (2001). *International Classification of Functioning, Disability and Health (ICF).* WHO. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health
Winser, S. J., Smith, C. M., Hale, L. A., Claydon, L. S. & Whitney, S. L. (2018). Clinical assessment of balance using BBS and SARAbal in cerebellar ataxia: Synthesis of findings of a psychometric property analysis. *Hong Kong Physiotherapy Journal, 38*(01), 53-61. https://doi.org/10.1142/S1013702518500063
Yabe, I., Matsushima, M., Soma, H., Basri, R., & Sasaki, H. (2008). Usefulness of the Scale for Assessment and Rating of Ataxia (SARA). *Journal of the Neurological Sciences, 266*(1–2), 164–166. https://doi.org/10.1016/j.jns.2007.09.021
YunBok, L., SeungHeon, A., & Gyu Chang, L. (2018). Clinical utility of the modified trunk impairment scale for stroke survivors. *Disability & Rehabilitation, 40*(10), 1200–1205. https://doi.org/10.1080/09638288.2017.1282990
Zhao, J., Chau, J. P. C., Zang, Y., Choi, K.C., He, R., Zhao, Y., Xiang, X., Li, Q. & Thompson, D. R. (2021). Psychometric properties of the Chinese version of the Trunk Impairment Scale in people with a stroke. *Health and Quality of Life Outcomes, 19*(1), 1-7. https://doi.org/10.1186%2Fs12955-021-01730-y
Zurawski, E., Behm, K., Dunlap, C., Koo, J., Ismail, F., Boulias, C., Reid, S. & Phadke, C. P. (2019). Interrater reliability of the Modified Ashworth Scale with standardized movement speeds: A pilot study. *Physiotherapy Canada, 71*(4), 348-354. https://doi.org/10.3138/ptc-2018-0086

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

0 **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. Biering-Sørensen 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 Somato-sensations
118. Van Lieshout Test Short Version
119. OR/79-118

0 **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"

1

162. "pain evaluation"
163. "pain measurement"
164. "rehabilitation assessment"
165. "rehabilitation measure"
166. "rehabilitation instrument"
167. "impairment measure"
168. "impairment assessment"
169. "impairment evaluation"
170. "Disability evaluation"
171."disability assessment"
172. "disability measure"
173. "Functional* Test*"
174. "Functional* Instrument*"
175. "Functional* Evaluation"
176. "Functional* Assessment*"
177. "Clinical Rating Scale"
178. "Rating scale"
179. "diagnostic tool*"
180. "Physical examination"
181. OR/120-180
182. 78 AND 119 AND 181
183. 24 AND 119 AND 181

2

Appendix B

Usability Rating Scale				
Criteria	*FEI Requirement*	*Rating*	*Score*	*Record Info/ Comments*
Test position	*Sitting, supine or prone lying*	*Sitting= 0* *Supine or prone lying or sitting = 1* *Standing or ambulation required = 2*		
Equipment Required	*Low cost, portable, user-friendly equipment. Or no equipment*	*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*		
Required Training	*1 – 2 days training for new impairment measures, which may include online training*	*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*		
Patient population	*Generic*	*Condition specific (only one patient group tested, excluding control group) = 1,* *Generic (more than one patient group tested) = 0*		
Data produced	*Interval or ratio scaled*	*Interval or Ratio scaled = 0* *Ordinal = 1* *No quantitative data produced = 2*		
Total Score			*/11*	

FEI – Fédération Équestre Internationale

Appendix C

<u>Potentially eligible impairment tools</u>

1. Ashworth Scale
2. Function in Sitting Test
3. Scale for Assessment and Rating of ataxia (SARA)
4. Trunk Impairment Scale
5. Tardieu Scal,
6. Hand-held Dynamometry
7. Motricity Index
8. Trunk Control Test
9. International Standards for Neurological Classification of Spinal Cord Injury
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
21. Berg Balance Scale
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.