Inter-rater agreement among physical therapists in using Neurologic Movement System Diagnosis (N-MSD)

Hoda Farhadi

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Sujata Pradhan, Chair

Patricia Scheets

Patricia Noritake Matsuda

Program Authorized to Offer Degree:

Rehabilitation Science

University of Washington

Abstract

Inter-rater agreement among physical therapists in using Neurologic Movement System Diagnosis (N-MSD)

Hoda Farhadi

Chair of the Supervisory Committee:
Sujata Pradhan
Department of Rehabilitation Medicine

Movement optimization was adopted by the American Physical Therapy Association (APTA) as the core of physical therapy practice in 2013. While Physical Therapists (PTs) observe, assess, diagnose, and treat movement impairments in clinical practice, evidence-based clinical practice guidelines are centered around neurological medical diagnoses rather than movement deficits. Therefore, in physical therapy practice, pathophysiological impairments and health conditions often guide assessment, treatment, and prognosis instead of the movement deficits.

To make movement the focus of the clinical decision-making process, several movement system diagnoses were developed. While some of the previously published frameworks were commonly used in clinical practice and educational settings, few studies have examined the reliability, validity, and clinical application of any of these systems. This dissertation study examined the agreement among PTs in utilizing the Neurologic Movement System Diagnosis (N-MSD) framework in individuals with various neurological health conditions.

The N-MSD includes frameworks for movement observation and analysis of key functional tasks as well as movement-based diagnoses. This body of work examined the agreement among PTs, between experienced and early-career PTs and against an expert considered as the gold standard. The findings demonstrated that experienced PTs with prior knowledge about the Movement System were in higher agreement with an expert when making movement system diagnoses. Moreover, PTs demonstrated greater agreement, a narrower range of scores, with assessments of static tasks as compared to their assessments of dynamic tasks. Among dynamic tasks however, assessment of transitional movements such as Sit to Stand and Step Up showed more scattered results across PTs compared to gait. Subsequently, agreement among PTs in selecting movement system diagnoses with key features related to dynamic tasks was lower as compared to other diagnoses. The collective results of these studies suggest that with proper training on N-MSD, physical therapists could come to moderate to strong agreement when assessing certain functional tasks and consensus on a primary movement diagnosis. These research findings also lay the groundwork for future research efforts, including development of a more uniform process for analyzing tasks with emphasis on dynamic tasks, psychometric evaluation of the new evaluation framework, and insight into the underlying clinical reasoning during the examination process and for a movement system diagnosis.

This research was conducted with the approval of the Human Subjects Division of the University of Washington.

IRB ID: STUDY00006824

Date of Approval: March 13, 2019

Plain Language Summary

Patients with various neurologic medical diagnoses visit physical therapists (PTs). Improving movement is the main focus of physical therapy. When assessing patients, it is critical for PTs to examine all body systems that contribute to movement issues. Along with clinical tests, PTs observe how individuals perform daily life tasks such as Sit to Stand, Step Up/Down, Walking, etc. Accurate analysis of these core tasks is essential in identifying the primary movement deficit that interferes with patients' function. Grouping and labeling patients based on movement-focused diagnoses can improve communication among rehabilitation providers. These diagnoses also help with assessment accuracy, clinical decision-making, treatment selection, and prognosis. The physical therapy profession has developed movement-focused assessment systems and diagnoses. However, the level of agreement among PTs in using these systems and diagnoses was not studied. To address this, we examined the agreement among PTs in using a classification system called Neurologic Movement System Diagnosis (N-MSD).

We examined patients with various neurological health conditions, video-recorded them, and asked PTs to watch the videos and assess the patient participants. PTs observed patients performing core functional tasks and then assigned movement-focused diagnoses to them. We aimed to find out if PTs agreed with an expert and with each other in assessing patients.

Results showed that overall, PTs were in moderate agreement with a gold standard expert in assigning diagnoses. PTs with more experience and familiarity with N-MSD had higher agreement with the expert compared to novice therapists. PTs who were involved in research on movement-focused frameworks showed the highest agreement. For task assessments, PTs tended to agree more with the expert and each other when assessing static tasks. The agreement level was lower for assessment of dynamic tasks. This finding was partially expected, as dynamic

tasks could magnify more movement problems compared to stationary tasks like Quiet Sitting and Standing. However, PTs agreed more when assessing Gait and Complex Gait, compared to scoring of Sit to Stand and Step Up tasks. This was likely because observation of gait is well tailored into physical therapy education programs. PTs are trained to systematically observe and assess complex walking patterns. There is less structure or emphasis on systematic assessment of transitional tasks.

This study highlights the importance of structured training for PTs in movement assessment. Future research should include a more diverse patient sample to represent all the N-MSD diagnoses. Future research could build upon the findings of this dissertation to establish reliability and validity of a more standardized and unified framework for observational task analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to everyone who supported me throughout this dissertation journey.

I am profoundly grateful to my committee chair, Dr. Sujata Pradhan whose patience and steadfast support carried me through this process. She guided me when I felt stuck, brought insight during uncertain moments, and believed in me even when my motivation wavered. I am also deeply thankful to Dr. Patty Scheets, who generously put trust in me throughout this project, stayed with me with patience and grace and dedicated countless hours to this project. To my entire committee, thank you for your continuous support and for cheering me on every step of the way.

A special acknowledgment goes to Dr. Sally McCoy for her tailored support and guidance, which were especially meaningful as an international student navigating this academic path. I would like to thank Dr. Debbie Kartin whose welcoming nature and mentorship nurtured my growth from the very beginning.

I am incredibly thankful to the faculty and staff of the Rehabilitation Medicine Department for making the department feel like a second home. A special thank you to Vickie Corrin for going above and beyond to provide support in ways that no one else could. Dr. Torey Gilbertson, for always being there for the PhD crew whenever we needed him. I am grateful for the warm support of the Physical Therapy Division faculty throughout my teaching journey. Your support made all the difference, and I had the best years of teaching alongside you all.

I would like to extend my sincerest thanks Dr. Mike Studer for his invaluable contribution to this project. His generosity in opening the doors of his clinic to me—allowing me to work with patients despite the challenges of the COVID-19 pandemic—was an incredible gift for which I am very grateful.

Thank you to my friends for their compassion and support. To everyone in our little PhD office—Caroline, Maryam, Helen, Anne, Lin-Ya, Cody, Kathryn, Majd, Beth, Jan, Misty and Thomas—your encouragement helped me through this journey and I am grateful for friendships that I know will last a lifetime. My friends who support me from afar, believed in me when I couldn't and reminded me of my strengths. The isolation of the pandemic led me to find a community through mutual love for music. I am thankful to the artists who shared part of their soul through their art, LT for being an incredible example of resilience and perseverance. I appreciate the friends I made because of him. The friendships and connections from afar kept me grounded and helped me stay connected to life.

Finally, to my family and relatives here—especially my lovely cousin's family, Ali, Hana, and Lianna—thank you for standing by me and supporting me through the ups and downs of this journey.

I am forever grateful to each and every one of you.

To my parents. The constant presence in my life. For their unconditional love.

To my sisters. For embracing me with endless love and compassion.

To my patients. Anywhere in this world. Gar be hame omr khish ba to bar aaram dami, haasel omr aan dam ast, baghi ayyam raft.

Chapter 1. Introduction

1.1 BACKGROUND

Identifying and providing the right intervention for patients is key in the healthcare systems to reduce unintended variation in practice. Selection of the appropriate intervention relies on demographic, clinical, and contextual factors for each individual patient. Health care providers often use diagnostic labels and classification systems to guide their clinical decision- making in determining the appropriate intervention plan.

Classification systems group patients with similar symptoms, clinical presentations, disease progression, and stages of health conditions in a systematic way (Edge & Compton, 2010). These systems guide the providers' clinical decision making by assisting them in choosing the appropriate treatment for patients with similar conditions. Medical diagnostic labels are also widely used in medicine. The function of these diagnostic labels is somewhat similar to classification systems. Diagnoses validate the presence of the disease or a condition, based on which, treatment algorithms are identified (Edge & Compton, 2010). Classification systems and diagnostic labels enable the use of standard terminology between health care (Rosenbaum et al., 2014). Moreover, use of standard classification systems enhances prediction of outcomes through uniformity of provided care for patients with similar medical conditions.

In traditional medical models, pathoanatomic or pathophysiologic terms are mainly used to guide clinical decision making and selecting treatments (McClure & Michener, 2015). Examples of medical diagnoses are tendinopathy, arthritis, capsulitis, etc. Diagnoses based on tissue-specific pathology or structure-based labels also point out the possible prognosis, especially if the label determines significant structural limitations (Cools & Michener, 2017; McClure & Michener, 2015).

In rehabilitation, the same medical, pathology-oriented diagnoses were historically used for patient assessment and management. However, depending on such structure-specific terminology introduces a number of challenges in the rehabilitation realm. In musculoskeletal physical therapy, many of the medical labels point out to the affected area without identifying the underlying cause of the impairments. Relying on the pathoanatomical consequence rather than the root cause of the problem makes it difficult to select the best plan of treatment for individuals (Cools & Michener, 2017; Ludewig et al., 2017; McClure & Michener, 2015). Additionally, some well-established diagnoses, such as rotator cuff syndrome, embrace a host of underlying movement impairments with very different physical therapy intervention processes. Easy communication among health care providers is a significant advantage of the pathoanatomic model, however, in many cases the diagnostic label primarily serves as a rule in/rule out mechanism to distinguish between surgical and non-surgical cases (Ludewig et al., 2017).

A similar limitation with labeling patients with medical diagnosis exists in neurologic physical therapy. Diagnoses such as cerebrovascular accident (CVA) or cerebral palsy (CP) only introduce the initial health condition without providing any information regarding functional ability or specific movement dysfunction of the individual (Van Sant, 2017a). Despite the fact that PTs organize their clinical decision making mostly around assessment of movement dysfunctions and activity limitations (Hedman et al., 2018; Saladin & Voight, 2017)

The discussion of recognizing PTs as diagnosticians of movement dysfunctions is not new. The debate could be traced back to as early as the 1970s with several visionaries in the profession contributing to this discussion through the decades (Coffin-Zadai, 2007). In 1983, the

American Physical Therapy Association (APTA) House of Delegates proposed a necessary focus on movement dysfunction as part of the definition of the profession (Coffin-Zadai, 2007). Despite the consensus on adopting the movement system approach with a focus on movement dysfunctions, the next steps in determining psychometric properties did not follow. Coffin-Zadai (2017) suggested some reasons for why the idea of developing diagnostic frameworks has not been at the front and center of the profession. Competing and pressing issues within the profession were given priority over movement-based diagnosis development due to complexity of the diagnostic process, lack of consensus on the uniform construct and language to be used for classification systems and the rapid evolution and impact of new knowledge.

The physical therapy profession has the treatment of movement issues at its core in the public eye (Saladin & Voight, 2017). In clinical practice, physical therapy interventions are developed to address certain movement dysfunctions. Meanwhile, in designing research studies that examine the effectiveness of these interventions, patients regularly are grouped based on their primary medical diagnosis. Hence, a group of individuals with the same medical condition, but heterogeneous and significantly different movement dysfunctions may receive a similar course of treatment due to health condition based clinical practice guidelines. This approach sometimes causes misinterpretation and misconception of the study outcomes (Van Sant, 2017a; Witten et al., 2023). Some have recommended that researchers interested in motor rehabilitation work across health conditions which may provide more meaningful outcomes and uptake by clinicians (Dobkin, 2009). Given that PTs organize their clinical decision making mostly around assessment of movement dysfunctions and activity limitations, a recent Clinical Practice Guideline (CPG) for improving walking function was developed across three health conditions, CVA, spinal cord injury, and brain injury (Hornby et al., 2020). Development and utilization of movement-specific labels would enhance this work by providing more specific guidance and advancing practice in neurologic physical therapy.

The use of movement-related terminology helps therapists to immediately build a mental picture of patients by focusing primarily on movement deficits (Scheets et al., 2007). Communication between rehabilitation providers through a set of clear and well-defined diagnostic labels improves the collaboration between therapists and third-party payers as well. Thus, systematically developed diagnostic labels, derived from clinical examination findings that describe movement dysfunctions are required (Hedman et al., 2018). Therefore, to advance the physical therapy profession, a paradigm shift in the classification and labeling process toward movement focused diagnosis was proposed (Ludewig et al., 2017; S. Sahrmann et al., 2017; S. A. Sahrmann, 2014).

As mentioned previously, the idea of grouping patients based on similar movement impairments is not new to PTs. About half a century ago, Brunnstrom developed a set of clinical tests to determine a patient's stage of recovery post-stroke (Brunnstrom, 1970). This treatment approach corresponded to the movement pattern dysfunction that a group of patients presented with. While the proposed techniques may be outdated, her practice of developing treatments for distinct movement patterns is still relevant today (Van Sant, 2017).

Simultaneous to the advances within the APTA House of Delegates in 1980, Sahrmann and associates, published within musculoskeletal physical therapy, the first version of (what later was known as) Movement System Impairment (MSI) Syndromes (S. Sahrmann, 2002; S. Sahrmann et al., 2017; Van Dillen et al., 1998). With its roots in the conceptual framework of kinesiopathologic model, the premise of the MSI system is that repetition of impaired movements and sustained mal alignments can cause pathoanatomical signs and symptoms. To restore ideal alignment and optimal function, correction of biomechanical or neuromuscular components of the impaired movement is essential (Caldwell et al., 2007; S. Sahrmann et al., 2017). The MSI syndromes were further developed for all body regions, covering the spine and major joints in upper and lower extremities. Numerous case studies examined the reliability and validity of the MSI systems throughout the decades. Several case reports also were published on implementation of MSI-based treatments for hip, knee, and low back pain, etc.

Similar to the efforts within musculoskeletal physical therapy, the Neurologic Movement System Diagnosis (N-MSD) developed by Scheets et al., was first introduced almost 2 decades ago (P. K. Scheets et al., 1999). To introduce the framework, Scheets and colleagues took a similar approach as the MSI syndromes, illustrating the application of the diagnoses in case examples with diverse neurological medical conditions (Scheets et al., 2007; P. K. Scheets et al., 1999; P. L. Scheets et al., 2015). One of the goals in developing this system was to be comprehensive in the patients categorized and their presenting problems including patients across health conditions, setting, and degree of activity limitation.

The N-MSD framework has evolved throughout the years, with periodic updates to the application manual and diagnosis tables. Not as many case reports were officially published compared to the uptake of MSI syndromes, but the application of the N-MSD was studied and disseminated through conference educational sessions, master classes and workshops and case report poster presentations. Additionally, efforts were put into standardization of the examination process of the N-MSD framework. While most of the tests within the N-MSD manual are well-established tests used in routine physical therapy assessments (e.g. range of motion, muscle strength, etc), both Motorneuron Response Assessment and Fractionated Movement tests developed as part of the project were studied and shown to be reliable and valid to use in neurologic patient populations (Scheets et al., 2007). However, to our knowledge there has not yet been a study that has investigated the reliability and validity of this entire classification system.

1.2 Summary Of Studies (aims and objectives)

The overall goal of this dissertation is to examine the inter-rater agreement among PTs using the N-MSD framework in patients with neurologic health conditions. The first study in this dissertation, "Inter-rater agreement between physical therapists and an expert in using Neurologic Movement System Diagnosis (N-MSD)", compares the agreement between PTs and an N-MSD expert in assessing patients with neurological conditions. The objective of this study is to explore the level of agreement between PTs and the expert on N-MSD label assignment and functional task observational analysis. The second study in this dissertation, "Inter-rater agreement among physical therapists in using Neurologic Movement System Diagnosis (N-MSD)", explores the level of agreement among PTs when using the N-MSD framework in patient assessment. The goal of this study is to examine the agreement levels for diagnostic label assignments and task analysis, among a group of therapists with varying degrees of previous N-MSD knowledge and neurologic clinical experience, when assessing patients with different neurologic health conditions. The third study in this dissertation, "Inter-rater agreement between experienced and early-career physical therapists in using Neurologic Movement System Diagnosis (N-MSD)", examines the agreement levels for diagnostic label assignments and task analysis in two groups of PTs, early-career and experienced, in using N-MSD when assessing patients with neurologic health conditions. The aim of this study is to determine if there was a difference in agreement levels between the two groups in using the N-MSD framework for assessing patients.

1.3 ORGANIZATION OF DISSERTATION

This dissertation follows the linked-studies format. Chapter one provides the background, knowledge gap in the pertinent literature and rationale for the research. Chapters two, three, and four are independent studies that will be prepared for submission to peer-reviewed journals. Each study is formatted to include the following sections: abstract, introduction, methods, results, discussion, conclusion, and references. Chapter five provides a summary of the dissertation, overarching conclusions, and ideas for future studies.

Chapter 2. Inter-rater agreement between physical therapists and an expert in using Neurologic Movement System Diagnosis (N-MSD)

2.1 INTRODUCTION

PTs assess and treat patients with movement impairments on a daily basis. In the physical therapy literature, classification systems were developed for various purposes such as education, research, and clinical practice within different physical therapy care settings. In the early decades of the profession, the traditional pathoanatomical medical frameworks were widely used, anchoring on the available science. However, by the turn of the century, leaders advocated for more evidence-based approaches to be used in educational and practical settings (Deutsch et al., 2022; L. P. Scheets et al., 2021).

In 2013, The American Physical Therapy Association (APTA) adopted movement optimization as the core of the physical therapy practice (*Vision Statement for the Physical Therapy Profession and Guiding Principles to Achieve the Vision*, 2013). To optimize movement, it is necessary to systematically identify the root of movement dysfunctions and target the underlying causes rather than signs and symptoms (*Vision Statement for the Physical Therapy Profession*, 2019). While the idea of movement-specific system labels is not new to our profession, there is a lack of rigorous studies that examine and support the reliability, validity, and clinical application of these systems.

In 2015, the Academy of Neurologic Physical Therapy (ANPT) appointed a Movement System (MS) Task Force to explore relevant extant classification systems for patients with neurological impairments. The group published a summary of their extensive literature review and recommendations regarding the essential elements of a movement system diagnosis framework in 2018 as a white paper (Hedman et al., 2018). The Task Force analyzed available movement-related classification systems and developed a recommendation framework as a foundation for future research in development and application of movement system diagnoses. Based on their recommendations for development and validation, the movement system should (1) be developed from a science-based theoretical framework, (2) emphasize movement examination and analysis through observation of core standardized daily life tasks, (3) comprise a set of movement system diagnoses where each diagnosis represents a cluster of movement observations and associated impairments, and (4) develop non-ambiguous labels for each movement system diagnosis.

Numerous tools, checklists or outcomes measures are used by PTs for observational movement analysis. While the reliability, validity and implementation of many of these tools were studied, they often focus on a specific movement or task (e.g. gait) in a specific patient population (e.g. patients with stroke). The observational analysis frameworks, however, use structured and systematic approaches to analyze movement across tasks and individuals. The Movement Continuum method described by Hedman et. al. is among the earliest examples of a structured framework for observational analysis of tasks (Hedman et al., 1996). They proposed common stages within a task from initial conditions to outcome like preparation, initiation, execution and termination that could be easily applied to any activity. Researchers built on this concept in following decades. In 2007, to enhance clinical relevancy of the Movement Continuum, Allen proposed a multidimensional model comprised of flexibility, strength, accuracy, speed, adaptability, endurance (Allen, 2007). A recent publication by ANPT MS taskforce proposed a structured framework of 11 movement constructs observable across six core functional tasks (Sitting, Sit to Stand, Standing, Walking, Step Up/Down, and Reach/Grasp/Manipulate) (Quinn et al., 2021). All of these structured frameworks aimed to enhance the understanding and management of movement dysfunction by using a unified framework of movement analysis. However, the reliability and implementation of their proposed constructs were not extensively studied.

Development of movement-specific assessment approaches and diagnostic labels followed a similar trajectory in physical therapy. In neurologic physical therapy, some of the traditional treatment approaches were designed to identify movement patterns or a cluster of clinical signs, with less focus on the primary neurologic medical conditions (e.g., indication for using neurodevelopment treatment (NDT) and proprioceptive neuromuscular facilitation (PNF) in patients with stroke and Cerebral palsy (CP). After the introduction of Movement System Impairment (MSI) syndromes within musculoskeletal physical therapy, a group of PTs created a MS framework for patients with neurologic health conditions by adopting a similar approach. The Neurologic Movement System Diagnosis (N-MSD) developed by Scheets et al., in the 1990's, was largely comprised of clinical guidelines developed and practiced in the authors' teaching clinics and hospitals for (treatment of) patients with neurological impairments (P. K. Scheets et al., 1999). Examination in this system consists of key tests for impairment in body function as well as analysis of critical functional tasks that are performed by patients.

The clinical tests used during the physical examination part of the N-MSD protocol are widely used in standard physical therapy examination sessions. These tests were designed to identify deficits in motor control, muscle tone (level of hyperexcitability), muscle strength (force generating capacity), nonequilibrium coordination, sensation, postural control, motion sensitivity, mental status, and joint range of motion. The psychometric properties of these clinical tests were previously established within the physical therapy literature.

The physical examination follows by assessment of key functional tasks including Quiet Sitting, Sit to Stand, Quiet Standing with feet hip's width apart and with feet together, Step Up (i.e., placing one foot on a step and returning it to the floor), Walking, walking while turning the head, stepping over obstacles, and walking forward and backward. These core functional tests were selected because they measure movement variables that are key to differentiate among MS problems. In an attempt to standardize the assessment of core functional tasks, the N-MSD adopted the Movement Continuum format for observational movement assessment (Hedman et al., 1996). The tasks were broken down into stages of movement: initiation, execution, and termination. The diagnostic element of the function-related tests is the observation of how a person accomplishes the activity, not the level of physical or verbal assistance needed. A checklist of possible movement deficits related to sequencing and timing of movement of body segments, body weight support, postural stability, and limb/trunk alignment are provided for each core task. During observational analysis of core tasks, PTs select the deficits (from this checklist) present with the patient.

Together, the examination components enable PTs to differentiate among movement system problems. The results from these tests and observational task analysis guide PTs in identifying the primary movement dysfunction. Aligned with the advantages of physical therapy-specific classification systems discussed before, the diagnostic labels in this system groups patients based on similarity in movement impairments rather than medical pathology or health condition.

Since the introduction of this system, the same research team published case studies and research articles to demonstrate the application of this classification system in a variety of patient populations with movement dysfunction resulting from neurologic health conditions (Scheets et al., 2007). Some of the salient features of the N-MSD are that this system was developed through years of working with various patients with neurologic conditions and could be used with more than one specific patient population. Additionally, the examination includes both body structure and function impairments as well as activity limitations. This is consistent with the model introduced by International Classification of Functioning, Disability and Health (ICF) (World Health Organization, 2001).

Considering the recommendations suggested by the ANPT MS Taskforce, the first step in the widespread implementation of MS diagnoses is to study the psychometric properties of extant systems. Additionally, while the face validity of the N-MSD was examined during the development and repeated clinical use, no controlled research study has examined the validity and reliability of the entire system (Hedman et al., 2018).

The purpose of this study was to examine the agreement in ratings by PTs and ratings from a Gold Standard expert (GS), in analyzing different tasks based on the N-MSD and assigning diagnostic labels in individuals with various neurological health conditions. We expect to see moderate/acceptable agreement between therapists and the GS.

2.2 Methods

The study was designed to take place in four overlapping phases. All procedures were approved by the University of Washington's Institutional Review Board. The overall goal was to establish the agreement between ratings by PTs and to those from the GS expert. The GS in this project was the primary person in the group of clinicians and researchers that developed the N-MSD. To meet the aims, the first step was to prepare and record patient videos and develop gold standard profiles for each patient. Next, recruited PTs were asked to review the patient profiles and assess and rate them using the N-MSD framework. The study roadmap is presented in Figure 2.1.

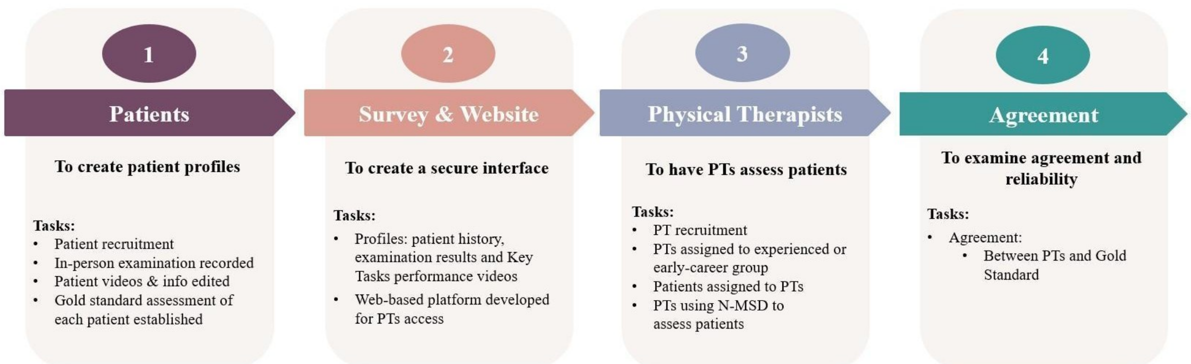


Figure 2.1 Study Roadmap

2.2.1 *Phase I: Patient Participants*

In preparation for patient recruitment, three pilot examination sessions were done to map out the camera setups and testing order. After patient screening, recruitment, and in-person examination sessions, profiles were created for the patients, reviewed, and edited for further use in the study. Details of these activities are outlined below.

Data collection forms; REDCap© platform

Data for this study were entered into and/or collected by instruments on a secure web platform, REDCap© (Research Electronic Data Capture) hosted at University of Washington (Harris et al., 2009, 2019). REDCap© is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. Surveys and instruments were designed and built in REDCap© by the Primary Investigator (PI) with assistance from administrative research organizations within the UW.

Clinic/lab environment and the study session

The setting for the in-person examination sessions was at the same physical therapy clinic that patients regularly visited. A couple of patients were seen at a research lab on the University of Washington campus. Two cameras were available to capture anterior-posterior and lateral views for all functional tasks.

Patient participants

Recruitment

The subjects were recruited from different sources such as physical therapy clinics, community centers, support groups, and through flyers placed in outpatient clinics across Oregon and Washington state. Potential subjects were contacted by email or phone calls initially. Subjects verbally agreed to be contacted prior to the initial contact by the PI. After being presented with a brief explanation about the study aims and upon their indication of interest, they were asked to fill out an online screening survey. The survey questions confirmed eligibility and asked for further demographic information. Adults with neurological conditions with the ability to walk 20 feet were eligible, regardless of level of assistance needed. Individuals who were not able to successfully ambulate 20 feet were not invited to participate in the study. The eligible individuals were contacted and scheduled for a one-time examination session.

Examination procedures

The examination sessions were conducted in a quiet environment with no interruption from other patients or researchers. The cameras were moved as needed during the sessions to have a frontal and lateral view of the patient at all times. Occasionally the room size did not allow for both views in which case the best angle was chosen by the examiner (PI).

The cameras were managed by the same examiner (PI) and were turned on at the beginning of the session after written consent was obtained. The sessions started with open-ended questions about the participant's medical history followed by routine physical exams. Patients were seated on a chair with no arm rests or at the edge of the bed with feet on the floor for most of the tests. Tests were done in a gradual manner based on the guidance provided within the N-MSD manuals, but occasional adjustments were made as needed. According to the N-MSD manual, the examiner was to provide as little assistance as possible during all of the functional tasks. In cases where participants were not able to complete a task, either limited by physical ability or fear of falling, the minimal amount of assistance required was given. Simpler functional tasks such as Quiet Sitting, Sit to Stand, and Quiet Standing were sometimes examined during impairment testing. Patients were offered water and rest breaks during the exam and the session concluded with more complex gait tasks. Participants were given $40 in the form of a Visa gift card at the end of the session.

Creating patient profiles

The patient data was initially curated and edited by the PI as described below. The results of the body structure and function tests were entered into REDCap© intake forms. Along with patient history and other relevant information, all the documented written data was compiled into a single portable document file (pdf). Next, footage from the two cameras were overlaid and adjusted for both picture frame and audio to prepare a single video clip of all functional tasks. For more clarity, solid and titled frames separated functional tasks from each other on the clip. The audio was regularly muted on videos for several reasons, mainly to protect patients' privacy or hide the conversation between patient and the examiner. If a verbal cue was given to the patient during the muted footage, closed captions were provided on the video to summarize what guidance was given to the participant. Given that only functional task performances were included in the video clip, the final edited video was usually less than 15 minutes in length (the shortest video being 5:27, and the average video length was 8:30). Due to a camera malfunction, the Step Up task in one video was recorded from posterior view only, and Gait task was only recorded for a few seconds.

A profile for each patient was created, consisting of the written data pdf file and core functional tasks video clip. The GS was given access to all participant profiles. Except for the initial profiles, the feedback was provided through asynchronized recorded sessions from the GS shared screen. On average, two profiles were reviewed in one sitting and returned with detailed feedback on the video frames of each functional task. In most cases, videos were edited to contain only relevant footage to the N-MSD framework. Suggested captions were added, and unnecessary minutes were removed. For some patient profiles, short clips of various additional tasks or tests, besides the core functional ones, were kept and labeled as extra clips. The decision was made jointly by PI and GS after reviewing each profile. While these extra minutes did not contain essential tasks from the MSD manual, there were diagnostic values to certain activities observed during that led to the decision of adding them to the patient examination videos.

Gold Standard assessment creation process

The GS completed the N-MSD assessment form while reviewing the data and chose a N-MSD diagnostic label at the end. In some cases, when a second diagnosis was applicable, it was recorded to be considered for further analysis. The result was saved as the GS assessment for each patient to be used later in the reliability examination phase of the study.

2.2.2 *Phase II: Website*

The goal of phase II was to a) develop an interactive online platform to upload and store all the patient profiles on, and b) create training modules on the N-MSD examination and diagnosis.

Website and The Patient Profile Pages

To securely share the patient data with PTs in the second phase, multiple options were considered. After considering the cost-benefit of developing an independent website, the research team decided to use a cloud-based collaborative platform for this study instead. Microsoft's SharePoint is an example of a cloud-based system that allows members to store, share, and manage data while providing various levels of security and access for each member. Additionally, creating a site through SharePoint does not require language-based coding. The site consisted of two major sections: patient profiles and training. Short videos of 2-3 minutes with explanatory narration were posted to orient users to the template of different parts of the website.

On each profile page, demographic information, history, and body function and structure test results were presented followed by the key functional tasks video and finally a link to the assessment form. There was no time limit to complete the assessment and participants could look at the patient information at any part of the assessment.

The Training Page

The training content was developed and reviewed by the GS and PI. The modules were in the format of PowerPoint© slides, embedded into the study training page. As PTs were expected to have different levels of clinical experience and MSD knowledge, the modules were introduced as a refresher and/or training material. While all PT participants were asked to complete the modules, each determined the pace and amount of time to spend on the material. We expected that on an average, participants would take 1-2 hours to complete the training, depending on their clinical experience, previous exposure to any MSD classification system, and established familiarity with N-MSD.

There were three training modules:

1. Introduction to MSD: This module covered an overview of the MS concept, the history and dialogue around the topic, and provided a brief list of other systems that were created and used throughout the decades in the physical therapy literature. The need for a Neurologic MS framework and early versions of N-MSD were also explored.

2. N-MSD diagnosis: A set of six slides were developed for each of the diagnoses, consisting of clinical presentation, considerations related to prognosis, treatment focus, and likely associated underlying pathophysiology. For each diagnosis, videos and/or photos of a case example were integrated. The last slide of a diagnosis had a knowledge checkpoint quiz. Questions on these knowledge checkpoints were written by the GS and intended to assess if participants were comfortable with the key elements and basic concepts relevant to the diagnosis. The checkpoints were not graded, and participants could see the correct answers immediately after completing the quiz.

3. The N-MSD assessment form: Adapted from the original N-MSD manual (published as Appendix to Scheets et al. 2014), a detailed assessment form was developed to include all the key functional tasks. Typical performance of each task was defined by the essential movement components for that task. Additionally, expected movement deficits for that task for each diagnosis was described. For transitional tasks, the observations were designated by phase of movement (initiation, execution, and termination). The number of possible movement deficits for each task varied, but items making up a task phase were separated by the phase label. In module 3 of the training page, the assessment form was available so that participants could familiarize themselves with it before beginning patient assessment.

Besides the training modules, supplemental documents of the original N-MSD publications were included on the training page. The PTs had access to the examination manual and diagnosis tables throughout the study.

2.2.3 *Phase III: Physical Therapist Participants*

The goal of phase III was to recruit PTs, have them review patient profiles, and assess patients using the N-MSD framework.

Recruitment

Individuals who worked in neurological settings and/or had 70% of their caseload focused on patients with neurologic health conditions were selected for this study.

The early-career PTs were sought through various methods such as contacting a) alumni of several Doctor of Physical Therapy (DPT) programs, b) neurologic residency programs for either current or former residents, and c) individuals suggested and referred by research team members and d) clinical instructors or other community PTs. Participants were deemed early career if they had less than five years of experience.

Experienced PT Participants

Individuals with years of experience in physical therapy practice with some knowledge of MS were sought out for this group of participants.

The experienced participants were recruited from multiple sources by the following methods: a) recommended by study team, b) through inquiry from former or current members of several ANPT task forces, c) by contacting directors of several neurologic residency programs, and d) recommended by PTs who participated and completed the study. Participants were deemed as experienced PTs if they had more than five years of clinical experience.

After the initial contact by the PI, all candidates were asked to complete a screening survey. Individuals were asked about their current working status, clinical workload and its patient population makeup, the length and history of their clinical practice, experience in neurologic settings, familiarity with the Movement Systems in general, any of the MSD frameworks and N-MSD in particular. Based on the answers provided, individuals were invited to participate in the study and assigned to either the early-career or experienced PT group.

Profile assignments to PTs

Two patients with different health conditions and N-MSD diagnoses were randomly assigned to each PT. Thus, each patient was seen by at least two PTs. In a few cases a PT reviewed more than two profiles, thus a patient was seen by more than four therapists total.

Access to the Website

Participants received website invitations upon inclusion criteria fulfillment. When accepting the invitation, they were added to the University of Washington (UW) cloud directory through a Microsoft email proxy. Assigned patient and PT participant identification (ID) numbers were emailed to each participant before website access. Participants were originally asked to complete the study within a 2–3-week period. Once the participation was underway, the PI checked on the progress periodically and sent reminders when needed. One participant only completed one out of two patient assessments.

2.2.4 *Phase IV: Examine agreement and reliability*

An estimation for sample size for both patients and PTs was suggested. This estimation was primarily based on similar previous studies of classification systems in the musculoskeletal literature

Data analysis

Data cleaning was done by the PI concurrently with the data collection. Eighteen subjects with a range of neurologic conditions were examined by the PI. After the recruitment and examination sessions, four patients were excluded from further stages of the study. Two of these individuals required more than 50% assistance (moderate to maximum assistance) for ambulation, one needed assistance from more than one person for more advanced skills such as Step Up and Gait. Another individual experienced mood and emotional disturbances during the first three functional tasks and the examination session had to end abruptly. Data from 14 participants were used in this study.

Reliability and agreement are both important issues when developing and testing new classification systems. Agreement is the degree to which measurements or ratings are identical (Kottner et al., 2011). Therefore, to examine the similarity between PT and the GS ratings in using the MSD framework, agreement estimates were considered appropriate. The type of variables involved in the analysis (categorical versus continuous) are other factors to consider when choosing estimates of agreement. To report agreement on a nominal variable, such as N-MSD labels, proportion and percentage agreement were reported. Agreement among raters on continuous measures, such as core task scores, range of agreement, or range of proportions of agreement were more meaningful to report.

Statistical Methods

When comparing the selected N-MSD diagnostic label, perfect agreement and near-perfect agreement were defined. If the first label chosen by a PT matched the first or second label chosen by the GS, the agreement was coded as perfect agreement. If a second label was selected by a PT and it matched the first or second label of the GS, the agreement was considered a near-perfect.

Examination of agreement in functional tasks assessment involved some data preparation. All individual items listed under a task on the assessment form, were binary coded into 1, if checked by the PTs as present, and 0 if left blank. The only exception was for the few items that were listed under the essential component phase. When reviewing the data, the research team decided not to include these few items at the top of each task sub-phases. Usually, absence or lack of a movement element is more significant for determining the movement issues. The mere presence of essential components is believed to not necessarily be of value for diagnosing movement system deficits and only be used as a checklist when assessing a task. Thus, these items were left out of further analysis. Ultimately, the binary score of all individual items were added up to create summary task variables. Thus, this summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. To examine agreement on task assessment between PTs and GS, mean, median, mode, standard deviation, and range of scores were reported for each functional task. Statistical analysis was done using the Statistical Package for the Social Sciences (SPSS) version 29. The assessment form data was transferred to the software.

2.3 Results

The purpose of this study was to examine the agreement in ratings between PTs and a GS expert in analyzing different tasks based on the N-MSD and assigning N-MSD diagnostic labels in individuals with various neurologic health conditions.

Fourteen patients with neurological conditions and 23 PTs were included in this study. Descriptive data for patients and PTs are summarized in table 2.1 and 2.2. Previous knowledge and exposure of PTs to topics related to MSDs are shown in Table 2.3. The majority of the PT participants (n=15) learned and followed the updates about neurology-specific MSD frameworks within the past decade, including the publications by the ANPT MSD Task Forces, presentations, and educational sessions at conferences.

Table 2.1 Descriptives of Physical Therapist Participants

Characteristics	Total	
	n = 23	%
Demographics		
*Employment Status **		
Full time PT job	18	78.3
Research, Teaching or Part time PT job	7	30.4
*Settings of practice ***		
Inpatient (Hospital)	8	35.6
Outpatient/Private Practice	6	26.1
Years in Practice		
5 years or less	12	52%
Neuro Case workload		
70% and more	18	78.3
Neurologic Clinical Specialist (NCS)	11	47.8

Note. PT = Physical Therapist. **Participants were allowed to choose more than one option. **Participants who worked only at one of the two settings.*

Table 2.2 Descriptives of Patient Participants

Measure	*n=14*	*%*
Demographics		
Male	6	43
Female	8	57
Age (M, SD)	62.72	17.43
Time since Diagnosis (M, SD)	13.14	9.31
Use of Assistive Device	5	36
Health Condition		
PD	5	36
MS	3	21
CVA	2	14
Mix	4	29
N-MSD Dx		
1.Movement Pattern Coordination Deficit (MPCD)	4	29
2.Force Production Deficit (FPD)	5	36
3.Fractionated Movement Deficit (FMD)	3	20
4.Postural Vertical Deficit (PVD)	0	0
5.Sensory Selection & Weighting Deficit (SSWD)	1	7
6.Sensory Detection Deficit (SDD)	0	0
7.Dysmetria (DM)	1	7
8.Hypokinesia (HK)	0	0

Note. N-MSD = Neurologic Movement System Diagnosis, CVA = Cerebrovascular Accident, PD= Parkinson's Disease, MS= Multiple Sclerosis.

Table 2.3 Descriptives of MSD knowledge of Physical Therapist participants

Measure	Total	
	n = 23	*%*
*MSD knowledge**		
Not at all	2	8.7
I have heard about the concept, but I don't know about any particular system	3	13
I am familiar with the concept, and have heard about couple of systems or diagnostic labels, but never used any in practice	5	21.7
I am familiar with the concept and examples, and have used some neuro systems/labels in my practice, occasionally as they fit	7	30.4
I am familiar with the concept and examples and use neuro systems/labels often in practice	6	26.1
*Introduction to MSD**		
I learned about the concept decades ago when it was first introduced (~1990's)	1	4.3
I learned about specific orthopedic systems in 2000's	1	4.3
I learned about the concept within the past decade	8	34.8
I have followed the publications and correspondences on MSD since 2015	7	30.4
I have read the 2017 APTA White Paper: "Movement System Diagnoses in Neurologic Physical Therapy"	15	65.2
I have attended presentations and talks at conferences, educational settings, at work, etc.	12	52.2
I have taken courses/attended workshops on a specific orthopedic system	2	8.7
I have taken courses/attended workshops on a specific neurological system	6	26.1
*N-MSD Knowledge**		
Are you familiar with this system?	15	65.2
I have attended presentations and talks at conferences, educational settings, at work, etc.	8	34.8
I have taken courses/attended workshops on this system	2	8.7
This system was taught in my DPT/Residency program	3	13
I learned about this system when the Academy of Neurologic Physical Therapy (ANPT) developed the Movement System Task Force in 2015	12	52.2
I have read the papers published by the Movement System Task Force in 2021	12	52.2
*Use of N-MSD in clinical practice**		
I don't currently use it	3	13
I use this system occasionally in my clinical practice	2	8.7
I use this system routinely in my clinical practice	7	30.4

I am teaching this content in a DPT/residency program	8	34.8
I am involved with research in this area	2	8.7

*Note. PT = Physical Therapist. *Participants were allowed to choose more than one option.*
***Some sections included PTs that did not select any of the listed items.*

To examine the agreement in ratings, each assessment done by a PT was compared against the assessment rated by the GS for the same patient. Then, a summary of the agreement in label assignment and task assessment followed.

Agreement in N-MSD label assignment:

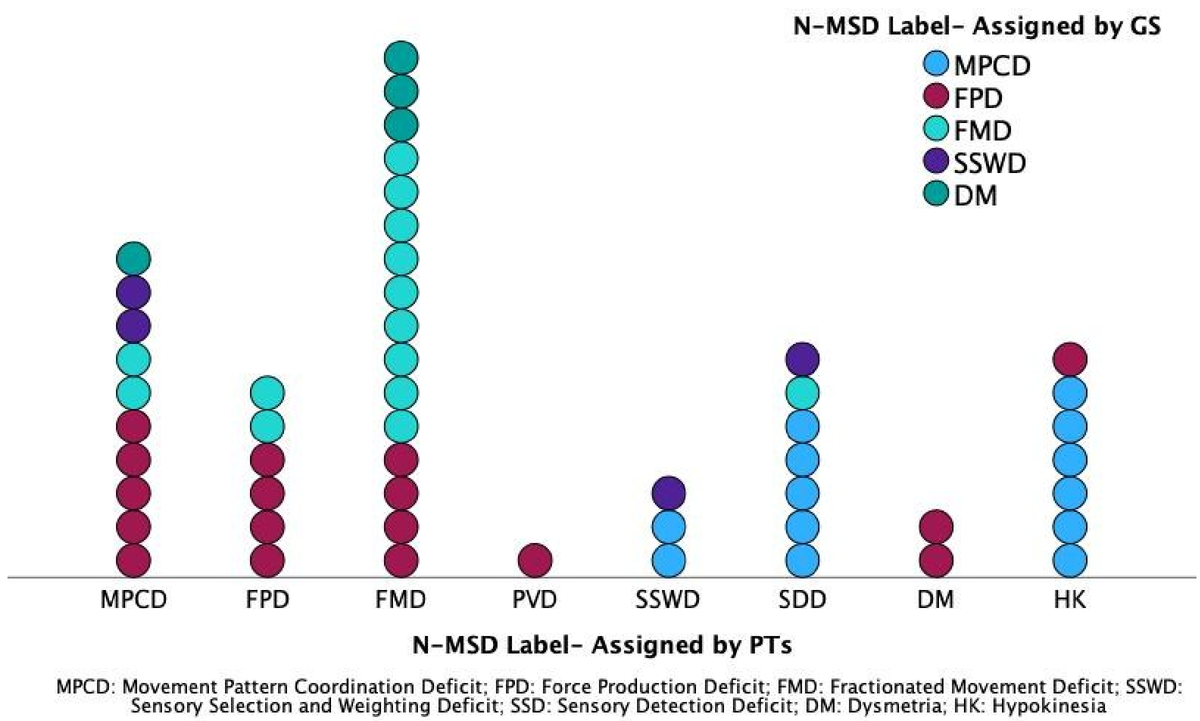
For the nominal variable of N-MSD label assignment, the proportion and percentage of perfect and near-perfect agreement estimates were shown in Table 4. In addition, assessments were grouped into similar N-MSD labels for further analysis. Overall, therapists and GS agreed in assigning N-MSD labels to patients in 40% of the assessments (21 out of 52 assessments). In further breakdown of results, the agreement between therapists and GS varied from 0 to 71% across different N-MSD diagnostic labels with Fractionated Movement Deficit (FMD) and Force Production Deficit (FPD) having the 1st and 2nd highest levels of agreement.

Table 2.4 Agreement Between Physical Therapists and Gold Standard in Assigning N-MSD Label

	Perfect Agreement		Near-Perfect Agreement		Overall Agreement	
	Proportion	Percentage	Proportion	Percentage	Proportion	Percentage
All therapists (n=23)	16/52	0.31	5/52	0.12	21/52	0.40
Agreement per N-MSD Diagnosis						
MPCD (patients=4)	0/13	0.00	0/13	0.00	0/13	0.00
FPD (patients=5)	6/17	0.35	4/17	0.24	10/17	0.59
FMD (patients=3)	9/14	0.64	1/14	0.07	10/14	0.71
SSWD (patients= 1)	1/4	0.25	0/4	0.00	1/4	0.25
DM (patients=1)	0/4	0.00	0/4	0.00	0/4	0.00

Note. Each PT assessed at least 2 patients. Each patient was seen by at least 2 PTs (except pt 12)
MPCD: Movement Pattern Coordination Deficit, FPD: Force Production Deficit, FMD: Fractionated Movement Deficit, SSWD: Sensory Selection and Weighting Deficit, DM: Dysmetria.

The label assignment spread is depicted in Graph 2.1. Each dot represents an assessment that was done by a PT. The x-axis displays the spread of the N-MSD labels that were chosen by PTs. The color of the dots shows the label that was assigned by the GS on each assessment. The highest level of agreement between PTs and the GS was in assigning the FMD label to patients (from 14 assessments that were labeled with FMD by the GS, 9 received the same diagnosis from PTs). In contrast, the least level of agreement between PTs and GS was when the GS labeled patients with MPCD. Of the 13 assessments labeled with MPCD by the GS, none received the same label from PTs. Instead, PTs were roughly divided between choosing HK and SDD diagnosis for the same cases.



Graph 2.1 N-MSD Label assignment by physical therapists and the GS

**This graph only shows the 1st diagnosis made by PTs and GS.*

Agreement in task assessment:

To examine the degree to which ratings were identical on task assessments, the task summary scores were used. As mentioned before, the task summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. To examine agreement on task assessment between PTs and the GS, the deviation score was calculated for all the assessments by subtracting the GS's score from PT's score for each task. Mean, median, mode, standard deviation, and range of the deviation scores were reported in Table 3. Besides the overall agreements, for further analysis, assessments on patients with a similar N-MSD label were clustered together.

Graph 2.2 shows the distribution of deviation scores for each key task across patients and PTs. The results of deviation scores, the PT assigned score difference from the GS score, are shown on the y-axis. Positive values indicate when PTs checked more items on the assessment forms in comparison to the GS. Negative values indicate when PTs selected fewer items than the GS when scoring a task performance. The boxes provide a visual summary of distribution, variability, and potential outliers in task assessments. The central line within each box represents the median value of deviation score, while the upper and lower edges of the box correspond to the interquartile range (IQR), spanning from the lower quartile (Q1) to the upper quartile (Q3). The whiskers extend to the smallest and largest observations within 1.5 times the IQR from the quartiles. Any deviation score beyond this range was considered an outlier and was plotted as an individual dot, with the adjacent number showing the patient number that the outlier assessment score represented.

The median value of most tasks was zero, except for Sit to Stand and Step Up that showed a median score of -1. This indicated that half of the deviation scores on these tasks were of values of -1 or smaller. The spread of data, as shown by the IQR, was smallest in Quiet Sitting (SD=0.82) and Quiet Standing (SD=1.17) and greatest in Sit to Stand (SD=2.64) and Step Up (SD=3.49) tasks, suggesting higher variability in deviation scores in these two tasks. The Step Up task also showed the widest range (-9 to 8) of deviation scores. Gait and Complex Gait tasks showed mean and median score deviations as low as the static tasks, however the standard deviation of the deviation scores were 1.64 and 2.41 respectively. The outlier deviation scores were seen in Quiet Sitting and Quiet Standing with feet together more than other tasks.

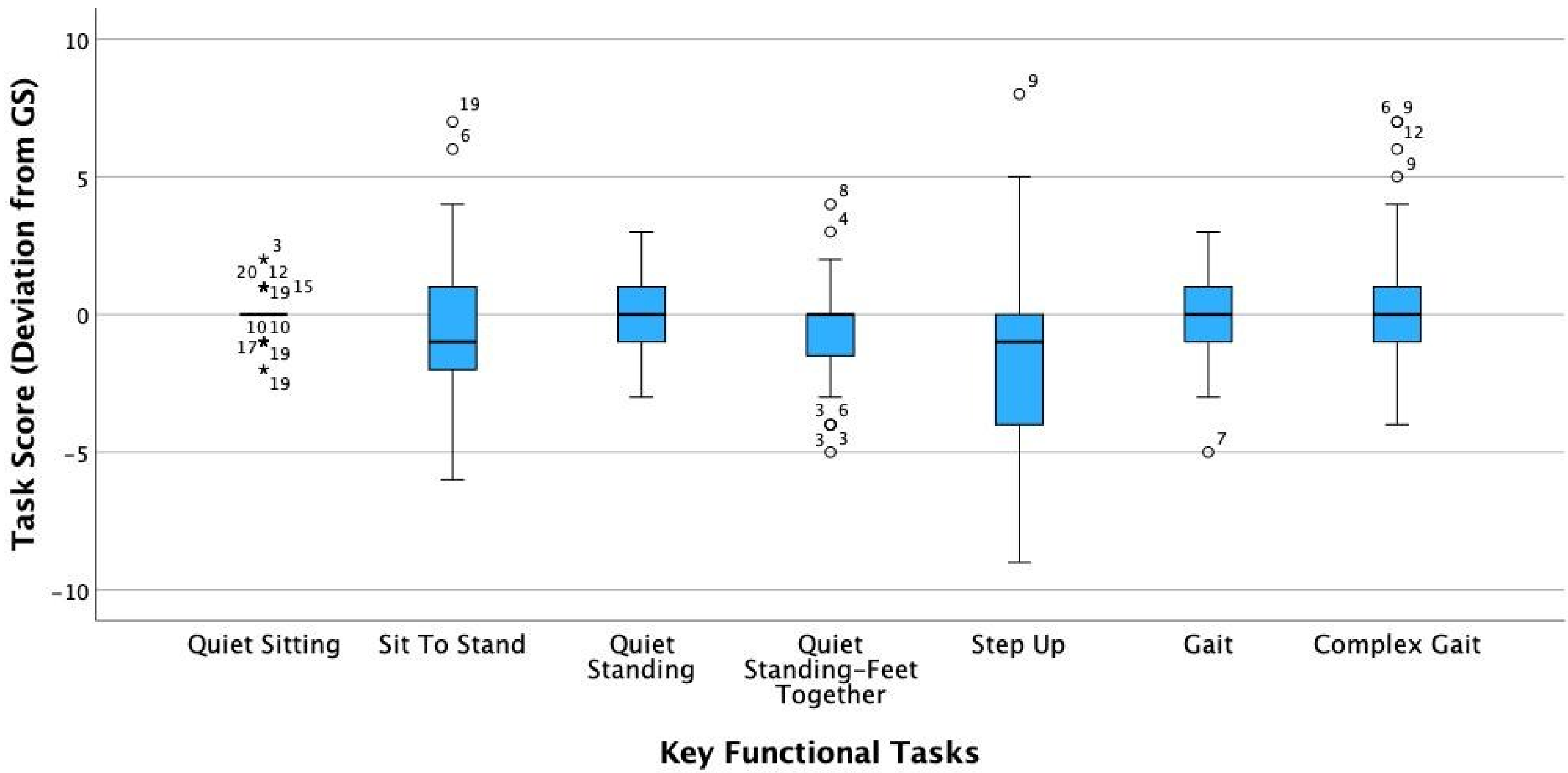
Across all tasks, deviation scores of patients number 3, 9, and 19 were marked as outliers on at least three tasks. Graph 2.3 provides a detailed representation of the score distribution across the transitional tasks, Sit to Stand and Step Up. The FPD and FMD groups account for the majority of the variability observed in the Sit to Stand task assessment. Similarly, assessment of the Step Up task in patients with FPD showed a great amount of variability among PT raters.

Table 2.5 Task Deviation Scores

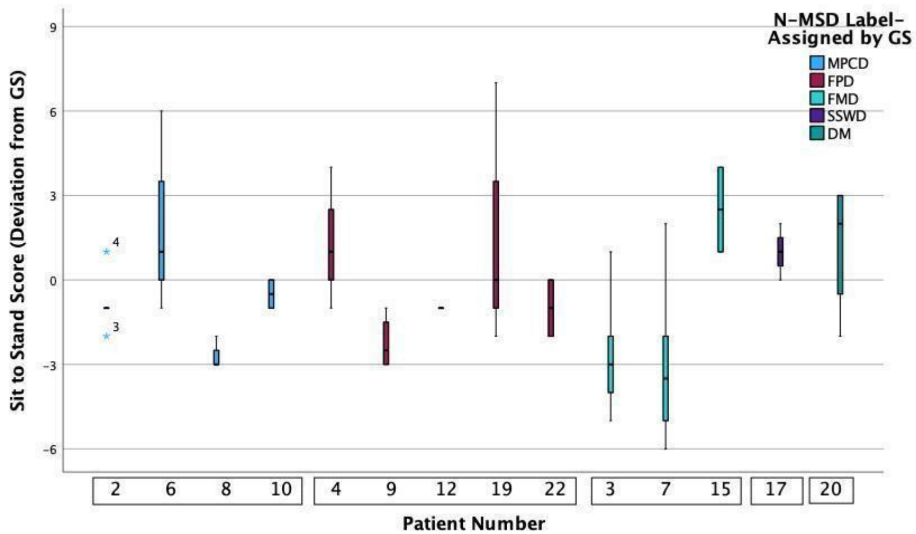
	Quiet Sitting						Sit to Stand					
	Mean	Median	Mode	SD	Minimum	Maximum	Mean	Median	Mode	SD	Minimum	Maximum
Overall Agreement (patients=14, PTs=23)	0.04	0	0	0.82	-2	2	-0.65	-1	-1	2.64	-6	7
Agreement per N-MSD Diagnosis												
MPCD (patients=4)	0	0	0	0.58	-1	1	-0.54	-1	-1	2.33	-3	6
FPD (patients=5)	-0.18	0	0	1.02	-2	2	-0.41	-1	-1	2.53	-3	7
FMD (patients=3)	0.21	0	0	0.89	-1	2	-2.07	-3	-3	2.97	-6	4
SSWD (patients= 1)	0.25	0	0	0.50	0	1	1	1	1	0.82	0	2
DM (patients=1)	0.25	0	0	0.50	0	1	1.25	2	3	2.36	-2	3

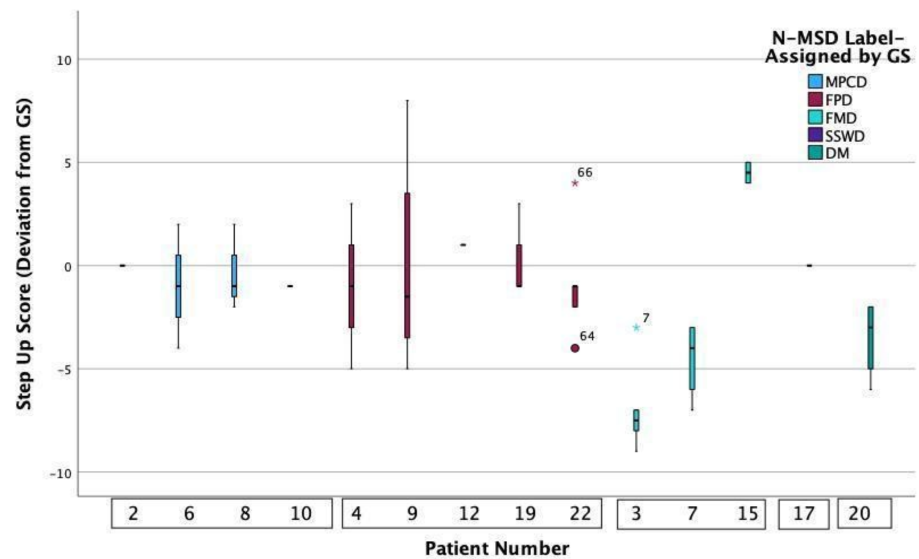
	Quiet Standing						Quiet Standing-Feet Together					
	Mean	Median	Mode	SD	Minimum	Maximum	Mean	Median	Mode	SD	Minimum	Maximum
Overall Agreement (patients=14, PTs=23)	-0.13	0	0	1.17	-3	3	-0.52	0	0	1.77	-5	4
Agreement per N-MSD Diagnosis												
MPCD (patients=4)	-0.62	0	0	0.77	-2	0	-0.15	0	0	1.95	-4	4
FPD (patients=5)	0.65	1	1	1.17	-2	3	-0.47	0	0	1.55	-3	3
FMD (patients=3)	-0.57	-0.5	-1a	1.22	-3	1	-1.29	-1	0	1.94	-5	2
SSWD (patients= 1)	0.25	0	0	0.50	0	1	0	-0.5	-1	1.41	-1	2
DM (patients=1)	-0.75	-0.5	0	0.96	-2	0	0.25	1	1	1.50	-2	1

	Step Up						Gait						Complex Gait					
	Mean	Median	Mode	SD	Minimum	Maximum	Mean	Median	Mode	SD	Minimum	Maximum	Mean	Median	Mode	SD	Minimum	Maximum
Overall Agreement (patients=14, PTs=23)	-1.65	-1	-1	3.49	-9	8	0.04	0	0	1.64	-5	3	0.33	0	0	2.41	-4	7
Agreement per N-MSD Diagnosis																		
MPCD (patients=4)	-0.46	0	0	1.56	-4	2	0.31	0	0	1.55	-2	3	1.46	1	-1	2.63	-2	7
FPD (patients=5)	-0.35	-1	-1	3.33	-5	8	0.06	0	0a	1.89	-3	3	0.71	0	0	2.69	-2	7
FMD (patients=3)	-4.29	-5	-7a	4.25	-9	5	-0.29	0	-1a	1.90	-5	3	-1	-1	0	1.80	-4	2
SSWD (patients= 1)	0	0	0	0.00	0	0	0	0	0	0.00	0	0	0.75	0.5	0	0.96	0	2
DM (patients=1)	-3.5	-3	-2	1.92	-6	-2	0.25	0.5	1	0.96	-1	1	-0.75	-0.5	0	0.96	-2	0



Graph 2.2 Deviation from the GS scores by each core tasks





Graph 2.3 Deviation from the GS scores

2.4 DISCUSSION

A prerequisite for using a MS classification framework is for therapists to be able to reliably classify patients into the suggested labels. The purpose of this study was to examine the agreement between PTs and a GS in using the N-MSD framework in patients with neurologic health conditions.

The label assignment analysis showed that therapists with different levels of clinical experience and previous familiarity with MS frameworks could assign the same diagnosis to patients as the GS, with moderate level of agreement (40%). However, the rate of agreement was not the same across PTs. Those with more years of clinical experience and previous familiarity with MS frameworks were in higher agreement with the GS. This is similar to the pattern found in reliability studies of other MS frameworks, such as the MSI classification system. In the context of the MSI, case studies to rigorously examine the validity and reliability of the system were initially conducted by the developers (Van Dillen et al., 1998). When therapists outside of the MSI developing team were recruited for a reliability study, it was shown that with appropriate training, therapists could reliably classify people with low back pain (LBP). Later, independent researchers examined the reliability of the region-specific MSI categories by using the system in clinical settings and patient populations (knee pain, hip pain, lower back pain, etc.) (Harris-Hayes & Van Dillen, 2009). The reliability estimates improved gradually in studies that included PTs outside of the research team but properly trained.

The training in this study included self-study of online, self-paced modules, with case examples for each diagnosis. While there were knowledge checkpoints integrated into the training modules to help with the learning process, there was no official monitoring of training completion by the PT participants. Therefore, the variance in agreement with the GS ratings among even more experienced PTs could point to the need for a more rigorous training prior to the use of N-MSD.

Analysis of label assignment between PTs and the GS also shows that in certain cases, therapists tend to choose a label that traditionally was paired with a medical condition. For example, for patients with Parkinson disease (PD) (n=5), PTs chose the hypokinesia label the most (7/14) when the GS did not find hypokinesia to be the primary or secondary movement deficit in any of these patients. This discrepancy could be due to relying on preconceived notions on how people move based on their medical diagnosis and not movement-specific signs and impairments relevant to physical therapy and outlined in the N-MSD framework.

Aside from the label assignment analysis, agreement in task assessments was also examined in this study. Analysis of the overall task scores showed that the most deviation from the GS was related to transitional tasks; Sit to Stand and Step Up. Half of the PTs selected at least one fewer deficits compared to the GS when assessing transitional tasks. Further analysis of the assessment on these two tasks showed greater variability among FPD and FMD labels. However, certain patients within these deficit groups were contributing the most to the score variability rather than all the patients within a group. Further investigation is needed to understand the underlying mechanisms contributing to these variability and differences from the GS. There were some variances in the deviation scores across patients and N-MSD labels in assessments of other static and dynamic tasks (Quiet Sitting, Standing and Gait). However, the variances did not point to a consistent pattern. In the N-MSD manual, key findings for each diagnosis are described. It is possible that the GS was observing specifically for the key findings associated with the diagnosis set as opposed to identifying all movement deficits present.

In further breakdown of the task assessments based on N-MSD labels, some groups demonstrated more deviations from the GS than the others. The assessments in the FPD group consisted of the widest deviation score ranges and biggest standard deviation estimates. While the label assignment agreement was the highest among FMD assessments (10/14, 71% agreement with GS), the fluctuation in task scoring was evident in this group. A further analysis showed that this wide spread of scores was produced by the assessments of two out of three patients in this group. The discrepancy of assessments in this group was the highest during Step Up and Sit to Stand tasks, respectively, and suggests that some patients are more difficult to diagnose than others.

Limitations

This study had some limitations. The N-MSD framework was designed for PTs to do a thorough in-person physical examination of patients before making diagnostic decisions. In this study, PT participants were required to make assessments based on written impairment findings and short videos of patients performing functional tasks. While the GS for examining movement dysfunction is 3-dimensional (3-D) movement analysis, use of 2 dimensional (2-D) videos of patient performance that could be also replicated in clinical settings have been established as a reliable and validated method. Several studies provided evidence for moderate to high reliability for observational gait assessment through use of videotapes (Brunnekreef et al., 2005; Ferrarello et al., 2013). A study showed that by watching visual 2- D videos of athletes performing single leg squat, novice and experienced PTs were able to classify lower extremity movement patterns in patients with musculoskeletal disorders (Harris-Hayes et al., 2014).

The examination process and commands used in the patient functional task assessment were not standardized in our study. The examiner relied on the N-MSD manual for required adjustments, facilitation, and commands during tasks performances, however, there were inconsistencies in execution of the tasks. Partially caused by space, time, and patient ability constraints, the examiner's discretion about modifications to the procedures may have introduced variations in task performances. Moreover, when reviewed by the GS, several assessments needed more in-depth investigation of findings that the examiner could have followed up with further testing.

The patient participants in this study were recruited from outpatient physical therapy clinics in the community. Hence, most of the participants were community-dwelling older adults capable of ambulating independently or with or without use of assistive devices or some assistance. This made the study sample limited in diversity in terms of individuals' function level, with no severe or acute case representation.

Gold standard reliability measures such as Kappa and Interclass Correlation Coefficients (ICCs) were not measured in this study due to the small sample size. Specific statistical approaches for assessment of reliability and agreement have been developed within different disciplines. Thus, there is no single statistical method to consider widely accepted to follow. It is also encouraged to report a combination of estimates when reporting the results to represent different statistical methods (Kottner et al., 2011). Nevertheless, the agreement proportions, percentages and range, although easily understandable, are limited to be extended for reliability purposes, as they do not adjust for agreements by chance. Other forms of reliability measures such as test-retest reliability are also recommended to be studied in future studies.

2.5 CONCLUSION

Focus on movement optimization through structured analysis of tasks and use of movement-based diagnoses is becoming more prevalent in physical therapy. The N-MSD is among the few organized MS frameworks that were developed and evolved through the years for assessment and management of patients with neurologic health conditions. The MS Task Forces within ANPT recommended examination of psychometric properties of MS frameworks as the future steps to take place in the profession.

The findings of this study showed a fair level of agreement between PTs and the GS in assignment of N-MSD labels as well as core task observational assessments when implementing N-MSD framework in assessment of patients with neurologic health conditions. The level of agreement varied across N-MSD labels. Additionally, in assessments done by PTs, some labels were noticeably paired with certain health conditions. Considering that the pairing was not evident in assessments done by the GS content expert, further training for PTs is required on some of the N-MSD labels.

Examination of reliability on key functional tasks showed that PTs tend to agree more with GS in assessment of static tasks. Dynamic and more complex tasks showed greater deviation from GS ratings among PTs, with Sit to Stand and Step Up consistently standing out with the highest level of variability in scores.

Future studies to further investigate this topic should deploy more rigorous statistical analytical approaches in examination of agreement. Adequate sample size and more unified training on the framework could enhance the robustness and applicability of the results.

Chapter 3. Inter-rater agreement among physical therapists in using Neurologic Movement System Diagnosis (N-MSD)

3.1 Introduction

PTs assess and treat patients with movement impairments on a daily basis. In the physical therapy literature, classification systems were developed for various purposes such as education, research, and clinical practice within different physical therapy care settings. In the early decades of the profession, the traditional pathoanatomical medical frameworks were widely used, anchoring on the available science. However, by the turn of the century, leaders advocated for more evidence-based approaches to be used in educational and practical settings (Deutsch et al., 2022; L. P. Scheets et al., 2021).

In 2013, The American Physical Therapy Association (APTA) adopted movement optimization as the core of the physical therapy practice (*Vision Statement for the Physical Therapy Profession and Guiding Principles to Achieve the Vision*, 2013). To optimize movement, it is necessary to systematically identify the root of movement dysfunctions and target the underlying causes rather than signs and symptoms (*Vision Statement for the Physical Therapy Profession*, 2019). While the idea of movement-specific system labels is not new to our profession, there is a lack of rigorous studies that examine and support the reliability, validity, and clinical application of these systems.

The importance of movement analysis in clinical practice and the challenges of achieving standardization were studied extensively (Ferrarello et al., 2013; Toro et al., 2003). Instrumented movement analysis through use of 3-dimensional (3-D) motion analysis, wearable devices. and non-wearable technologies are considered Gold Standard of gait analysis, as they provide reliable and accurate information (Hulleck et al., 2022; Petraglia et al., 2019). However, observational movement analysis remains the most popular method in clinical practice due to affordability and simplicity (Ferrarello et al., 2013; Gor-García-Fogeda et al., 2016; Ridao-Fernández et al., 2019). Therefore, observational movement analysis is primarily taught in physical therapy education programs.

Several tools, checklists and outcomes measures for observational movement analysis were developed and examined for psychometric properties in physical therapy literature. Among the most well established and early examples of gait analysis is the Observational Gait Analysis (OGA) developed by the Professional Staff Association of Rancho Los Amigos Medical Center (*Observational Gait Analysis Handbook*, 1989). The interrater reliability among PTs in rating binary items on OGA checklist showed low to moderate agreement (Eastlack et al., 1991). The outcome measures and scales developed after OGA, either focused on certain body parts during a task or were examined for their reliability and validity within specific patient populations. For example, a systematic review on observational scales in patients with gait disorders found the Edinburgh Visual Gait Score (EVGS) among the most comprehensive gait scoring systems that demonstrated a high methodological quality in assessing the gait of children with CP (Ridao-Fernández et al., 2019). The G.A.I.T. is a comprehensive 31-item measure divided into sections evaluating the upper and lower limb and trunk during the stance and swing phases of gait. It was described as having good to excellent reliability and sensitivity to change in patients with chronic stroke (Gor-García-Fogeda et al., 2016; Ridao-Fernández et al., 2019). The Tinetti Performance-Oriented Mobility Assessment (POMA) comprises a gait subscale with 10 items assessing various aspects of gait coordination. For patients with PD, POMA demonstrated predictive capability for falls and mortality, as well as intra- and interrater reliability (Gor-García-Fogeda et al., 2016). Reviews of these scales have highlighted varying levels of reliability and validity depending on the scale, the population studied, and the experience level of the raters. Therefore, structured assessment tools and standardized training are required for consistency in observational gait analysis.

A different approach to observational movement analysis, rooted in theoretical knowledge of motor control and rehabilitation science, was first proposed by Hedman et al. in 1998. Rather than describing movement pattern deviations, the authors highlighted the use of a structured and systematic framework, Movement Continuum, to analyze movement across tasks and individuals. Common stages within a task from initial conditions to outcome (preparation, initiation, execution and termination) could be easily applied to any activity. Researchers built on this concept in following decades. In 2007, to enhance clinical relevancy of the Movement Continuum, Allen proposed a multidimensional model comprised of 6 dimensions; flexibility, strength, accuracy, speed, adaptability, endurance (Allen, 2007)**.** A more recent work, adopting a MS approach, accentuated the four elements essential to all movement: motion, force, energy, and control. While these elements overlap and interact, they can each be examined and tested with some degree of specificity within any activity. The structured framework could enhance development of hypotheses about movement dysfunction and therefore the clinical decision making necessary for observational movement analysis. Furthermore, the model's application in sports physical therapy practice and education demonstrates its versatility and relevance across various clinical settings, underscoring its potential to standardize movement analysis and improve patient outcomes. A recent publication by a ANPT taskforce proposed a structured framework of 11 movement constructs observable across six core functional tasks (Sitting, Sit to Stand, Standing, Walking, Step Up/Down, and Reach/Grasp/Manipulate) (Quinn et al., 2021).

All of these structured frameworks aimed to enhance the understanding and management of movement dysfunction by using a unified framework of movement analysis. However, the reliability and implementation of their proposed constructs were not extensively studied.

In 2018, the appointed MS Task Force by the ANPT published a white paper on essential elements of a movement system diagnosis framework (Hedman et al., 2018). The Task Force analyzed available movement-related classification systems and developed a recommendation framework as a foundation for future research in development and application of movement system diagnoses. Based on their recommendations for development and validation, the movement system should (1) be developed from a science-based theoretical framework, (2) emphasize movement examination and analysis through observation of core standardized daily life tasks, (3) comprise a set of movement system diagnoses where each diagnosis represents a cluster of movement observations and associated impairments, and (4) develop non-ambiguous labels for each movement system diagnosis.

The N-MSD by Scheets et al. is among examples of available MS frameworks. First published in the 1990's, this system was developed through years of working with various patients with neurologic conditions and therefore could be used with more than one specific patient population. Since the introduction of this system, the developers published case studies and research articles to demonstrate the application of this classification system in a variety of patient populations with movement dysfunction resulting from neurologic health conditions (Scheets et al., 2015).

The N-MSD, an example of a structured and well developed framework for MS approach, consists of key tests for impairment in body function as well as observational analysis of critical functional tasks that are performed by patients. The clinical tests used during the physical examination part of the N-MSD protocol are widely used in standard physical therapy examination sessions. These tests were designed to identify deficits in motor control, muscle tone (level of hyperexcitability), muscle strength (force generating capacity), nonequilibrium coordination, sensation, postural control, motion sensitivity, mental status, and joint range of motion. The psychometric properties of these clinical tests were previously established within the physical therapy literature. The physical examination follows by assessment of key functional tasks including Quiet Sitting, Sit to Stand, Quiet Standing with feet hip's width apart and with feet together, Step Up (i.e., placing one foot on a step and returning it to the floor), Walking, walking while turning the head, stepping over obstacles, and walking forward and backward. These core functional tests were selected because they measure movement variables that are key to differentiate among MS problems. In an attempt to standardize the assessment of core functional tasks, the N-MSD adopted the Movement Continuum format for observational movement assessment (Hedman et al., 1996). The tasks were broken down into stages of movement: initiation, execution, and termination. The diagnostic element of the function-related tests is the observation of how a person accomplishes the activity, not the level of physical or verbal assistance needed. A checklist of possible movement deficits related to sequencing and timing of movement of body segments, body weight support, postural stability, and limb/trunk alignment are provided for various phases among each core task. During observational analysis of core tasks, PTs select the deficits (from this checklist) present with the patient.

Together, the examination components enable the PTs to differentiate among movement system problems. The results from these tests and observational task analysis guide PTs in identifying the primary movement dysfunction. Aligned with the advantages of physical therapy-specific classification systems discussed before, the diagnostic labels in this system groups patients based on similarity in movement impairments rather than medical pathology or health condition.

Considering the recommendations suggested by the ANPT MS Taskforce, the first step in the widespread implementation of MS diagnoses is to study the psychometric properties of extant systems. Additionally, while the face validity of the N-MSD was examined during the development and repeated clinical use, no controlled research study has examined the validity and reliability of the entire system (Hedman et al., 2018). Therefore, the purpose of this study was to examine the agreement in ratings among PTs, in analyzing different tasks based on the N-MSD and assigning diagnostic labels in individuals with various neurologic health conditions. We expect to see moderate/acceptable agreement among PTs across patients.

3.2 METHODS

The study was designed to take place in four overlapping phases. All procedures were approved by the University of Washington's Institutional Review Board. The overall goal was to establish the agreement between ratings by PTs and to those from the GS expert. The GS in this project was the primary person in the group of clinicians and researchers that developed the N-MSD. To meet the aims, the first step was to prepare and record patient videos and develop gold standard profiles for each patient. Next, recruited PTs were asked to review the patient profiles and assess and rate them using the N-MSD framework. The study roadmap is presented in Figure 3.1.

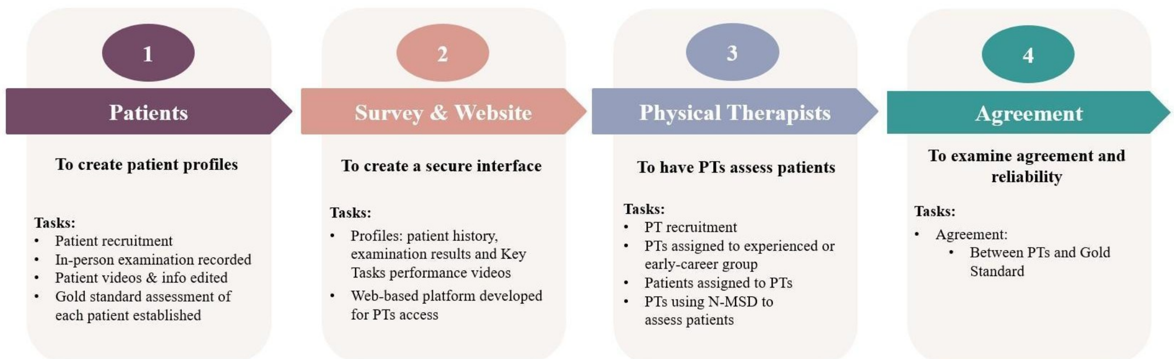


Figure 3.1 Study Roadmap

3.2.1 *Phase I: Patient Participants*

In preparation for patient recruitment, three pilot examination sessions were done to map out the camera setups and testing order. After patient screening, recruitment, and in-person examination sessions, profiles were created for the patients, reviewed, and edited for further use in the study. Details of these activities are outlined below.

Data collection forms; REDCap© platform

Data for this study were entered into and/or collected by instruments on a secure web platform, REDCap© (Research Electronic Data Capture) hosted at University of Washington (Harris et al., 2009, 2019). REDCap© is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. Surveys and instruments were designed and built in REDCap© by the Primary Investigator (PI) with assistance from administrative research organizations within the UW.

Clinic/lab environment and the study session

The setting for the in-person examination sessions was at the same physical therapy clinic that patients regularly visited. A couple of patients were seen at a research lab on the University of Washington campus. Two cameras were available to capture anterior-posterior and lateral views for all functional tasks.

Patient participants

Recruitment

The subjects were recruited from different sources such as physical therapy clinics, community centers, support groups, and through flyers placed in outpatient clinics across Oregon and Washington state. Potential subjects were contacted by email or phone calls initially. Subjects verbally agreed to be contacted prior to the initial contact by the PI. After being presented with a brief explanation about the study aims and upon their indication of interest, they were asked to fill out an online screening survey. The survey questions confirmed eligibility and asked for further demographic information. Adults with neurological conditions with the ability to walk 20 feet were eligible, regardless of level of assistance needed. Individuals who were not able to successfully ambulate 20 feet were not invited to participate in the study. The eligible individuals were contacted and scheduled for a one-time examination session.

Examination procedures

The examination sessions were conducted in a quiet environment with no interruption from other patients or researchers. The cameras were moved as needed during the sessions to have a frontal and lateral view of the patient at all times. Occasionally the room size did not allow for both views in which case the best angle was chosen by the examiner (PI).

The cameras were managed by the same examiner (PI) and were turned on at the beginning of the session after written consent was obtained. The sessions started with open-ended questions about the participant's medical history followed by routine physical exams. Patients were seated on a chair with no arm rests or at the edge of the bed with feet on the floor for most of the tests. Tests were done in a gradual manner based on the guidance provided within the N-MSD manuals, but occasional adjustments were made as needed. According to the N-MSD manual, the examiner was to provide as little assistance as possible during all of the functional tasks. In cases where participants were not able to complete a task, either limited by physical ability or fear of falling, the minimal amount of assistance required was given. Simpler functional tasks such as Quiet Sitting, Sit to Stand, and Quiet Standing were sometimes examined during impairment testing. Patients were offered water and rest breaks during the exam and the session concluded with more complex gait tasks. Participants were given $40 in the form of a Visa gift card at the end of the session.

Creating patient profiles

The patient data was initially curated and edited by the PI as described below. The results of the body structure and function tests were entered into REDCap© intake forms. Along with patient history and other relevant information, all the documented written data was compiled into a single portable document file (pdf). Next, footage from the two cameras were overlaid and adjusted for both picture frame and audio to prepare a single video clip of all functional tasks. For more clarity, solid and titled frames separated functional tasks from each other on the clip. The audio was regularly muted on videos for several reasons, mainly to protect patients' privacy or hide the conversation between patient and the examiner. If a verbal cue was given to the patient during the muted footage, closed captions were provided on the video to summarize what guidance was given to the participant. Given that only functional task performances were included in the video clip, the final edited video was usually less than 15 minutes in length (the shortest video being 5:27, and the average video length was 8:30). Due to a camera malfunction, the Step Up task in one video was recorded from posterior view only, and gait task was only recorded for a few seconds.

A profile for each patient was created, consisting of the written data pdf file and core functional tasks video clip. The GS was given access to all participant profiles. Except for the initial profiles, the feedback was provided through asynchronized recorded sessions from the GS shared screen. On average, two profiles were reviewed in one sitting and returned with detailed feedback on the video frames of each functional task. In most cases, videos were edited to contain only relevant footage to the N-MSD framework. Suggested captions were added, and unnecessary minutes were removed. For some patient profiles, short clips of various additional tasks or tests, besides the core functional ones, were kept and labeled as extra clips. The decision was made jointly by PI and GS after reviewing each profile. While these extra minutes did not contain essential tasks from the MSD manual, there were diagnostic values to certain activities observed during that led to the decision of adding them to the patient examination videos.

Gold Standard assessment creation process

The GS completed the N-MSD assessment form while reviewing the data and chose a N-MSD diagnostic label at the end. In some cases, when a second diagnosis was applicable, it was recorded to be considered for further analysis. The result was saved as the GS assessment for each patient to be used later in the reliability examination phase of the study.

3.2.2 *Phase II: Website*

The goal of phase II was to a) develop an interactive online platform to upload and store all the patient profiles on, and b) create training modules on the N-MSD examination and diagnosis.

Website and The Patient Profile Pages

To securely share the patient data with PTs in the second phase, multiple options were considered. After considering the cost-benefit of developing an independent website, the research team decided to use a cloud-based collaborative platform for this study instead. Microsoft's SharePoint is an example of a cloud-based system that allows members to store, share, and manage data while providing various levels of security and access for each member. Additionally, creating a site through SharePoint does not require language-based coding. The site consisted of two major sections: patient profiles and training. Short videos of 2-3 minutes with explanatory narration were posted to orient users to the template of different parts of the website.

On each profile page, demographic information, history, and body function and structure test results were presented followed by the key functional tasks video and finally a link to the assessment form. There was no time limit to complete the assessment and participants could look at the patient information at any part of the assessment.

The Training Page

The training content was developed and reviewed by the GS and PI. The modules were in the format of PowerPoint© slides, embedded into the study training page. As PTs were expected to have different levels of clinical experience and MSD knowledge, the modules were introduced as a refresher and/or training material. While all PT participants were asked to complete the modules, each determined the pace and amount of time to spend on the material. We expected that on an average, participants would take 1-2 hours to complete the training, depending on their clinical experience, previous exposure to any MSD classification system, and established familiarity with N-MSD.

There were three training modules:

1. Introduction to MSD: This module covered an overview of the Movement System (MS) concept, the history and dialogue around the topic, and provided a brief list of other systems that were created and used throughout the decades in the physical therapy literature. The need for a Neurologic MS framework and early versions of N-MSD were also explored.

2. N-MSD diagnosis: A set of six slides were developed for each of the diagnoses, consisting of clinical presentation, considerations related to prognosis, treatment focus, and likely associated underlying pathophysiology. For each diagnosis, videos and/or photos of a case example were integrated. The last slide of a diagnosis had a knowledge checkpoint quiz. Questions on these knowledge checkpoints were written by the GS and intended to assess if participants were comfortable with the key elements and basic concepts relevant to the diagnosis. The checkpoints were not graded, and participants could see the correct answers immediately after completing the quiz.

3. The N-MSD assessment form: Adapted from the original N-MSD manual (published as Appendix to Scheets et al. 2014), a detailed assessment form was developed to include all the key functional tasks. Typical performance of each task was defined by the essential movement components for that task. Additionally, expected movement deficits for that task for each diagnosis was described. For transitional tasks, the observations were designated by phase of movement (initiation, execution, and termination). The number of possible movement deficits for each task varied, but items making up a task phase were separated by the phase label. In module 3 of the training page, the assessment form was available so that participants could familiarize themselves with it before beginning patient assessment.

Besides the training modules, supplemental documents of the original N-MSD publications were included on the training page. The PTs had access to the examination manual and diagnosis tables throughout the study.

3.2.3 *Phase III: Physical Therapist Participants*

The goal of phase III was to recruit PTs, have them review patient profiles, and assess patients using the N-MSD framework.

Recruitment

Individuals who worked in neurological settings and/or had 70% of their caseload focused on patients with neurologic health conditions were selected for this study.

The early-career PTs were sought through various methods such as contacting a) alumni of several Doctor of Physical Therapy (DPT) programs, b) neurologic residency programs for either current or former residents, and c) individuals suggested and referred by research team members and d) clinical instructors or other community PTs. Participants were deemed early career if they had less than five years of experience.

Experienced PT Participants

Individuals with years of experience in physical therapy practice with some knowledge of Movement Systems were sought out for this group of participants.

The experienced participants were recruited from multiple sources by the following methods: a) recommended by study team, b) through inquiry from former or current members of several ANPT task forces, c) by contacting directors of several neurologic residency programs, and d) recommended by PTs who participated and completed the study. Participants were deemed as experienced PTs if they had more than five years of clinical experience.

After the initial contact by the PI, all candidates were asked to complete a screening survey. Individuals were asked about their current working status, clinical workload and its patient population makeup, the length and history of their clinical practice, experience in neurologic settings, familiarity with the Movement Systems in general, any of the MSD frameworks and N-MSD in particular. Based on the answers provided, individuals were invited to participate in the study and assigned to either the early-career or experienced PT group.

Profile assignments to PTs

Two patients with different health conditions and N-MSD diagnoses were randomly assigned to each PT. Thus, each patient was seen by at least two PTs. In a few cases a PT reviewed more than two profiles, thus a patient was seen by more than four therapists total.

Access to the Website

Participants received website invitations upon inclusion criteria fulfillment. When accepting the invitation, they were added to the University of Washington (UW) cloud directory through a Microsoft email proxy. Assigned patient and PT participant identification (ID) numbers were emailed to each participant before website access. Participants were originally asked to complete the study within a 2-3 week period. Once the participation was underway, the PI checked on the progress periodically and sent reminders when needed. One participant only completed one out of two patient assessments.

3.2.4 *Phase IV: Examine agreement and reliability*

An estimation for sample size for both patients and PTs was suggested. This estimation was primarily based on similar previous studies of classification systems in the musculoskeletal literature

Data analysis

Data cleaning was done by the PI concurrently with the data collection. Eighteen subjects with a range of neurologic conditions were examined by the PI. After the recruitment and examination sessions, four patients were excluded from further stages of the study. Two of these individuals required more than 50% assistance (moderate to maximum assistance) for ambulation, one needed assistance from more than one person for more advanced skills such as Step Up and Gait. Another individual experienced mood and emotional disturbances during the first three functional tasks and the examination session had to end abruptly. Data from 14 participants were used in this study.

Reliability and agreement are both important issues when developing and testing new classification systems. Agreement is the degree to which measurements or ratings are identical (Kottner et al., 2011). Therefore, to examine the similarity between PT and the GS ratings in using the MSD framework, agreement estimates were considered appropriate. The type of variables involved in the analysis (categorical versus continuous) are other factors to consider when choosing estimates of agreement. To report agreement on a nominal variable, such as N-MSD labels, proportion and percentage agreement were reported. Agreement among raters on continuous measures, such as core task scores, range of agreement, or range of proportions of agreement were more meaningful to report.

Statistical Methods

To compare the selected N-MSD diagnostic label among PTs all the assessments done on a patient were piled up. Next, agreement was calculated across patients with the same health condition and movement deficit. To account for the different number of PTs who evaluated each profile and therefore various number of assessments on each patient, a weighted mean was calculated for each category.

Examination of agreement in functional tasks assessment involved some data preparation. All individual items listed under a task on the assessment form, were binary coded into 1, if checked by the PTs as present, and 0 if left blank. The only exception was for the few items that were listed under the essential component phase. When reviewing the data, the research team decided not to include these few items at the top of each task sub-phases. Usually, absence or lack of a movement element is more significant for determining the movement issues. The mere presence of essential components is believed to not necessarily be of value for diagnosing movement system deficits and only be used as a checklist when assessing a task. Thus, these items were left out of further analysis. Ultimately, the binary score of all individual items were added up to create summary task variables. Thus, this summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. To examine agreement on task assessment among PTs mean, median, mode, standard deviation, and range of scores were reported for each functional task. Statistical analysis was done using the Statistical Package for the Social Sciences (SPSS) version 29. The assessment form data was transferred to the software.

3.3 Results

The purpose of this study was to examine the agreement in ratings among PTs in analyzing different tasks based on the N-MSD framework and assigning N-MSD diagnostic labels in individuals with various neurological health conditions.

Fourteen patients with neurologic conditions and 23 physical therapists were included in this study. Descriptive data for patients and PTs are summarized in Table 3.1 and 3.2. Previous knowledge and exposure of PTs to topics related to MSDs are shown in Table 3.3. Majority of the PT participants (n=15) learned and followed the developments in the neurology-specific MSD frameworks within the past decade, including the publications by the ANPT MSD Task Forces, presentations, and educational sessions at conferences, etc.

Table 3.1 Descriptives of Physical Therapist Participants

Characteristics	Total	
	n = 23	*%*
Demographics		
*Employment Status **		
Full time PT job	18	78.3
Research, Teaching or Part time PT job	7	30.4
*Settings of practice ***		
Inpatient (Hospital)	8	35.6
Outpatient/Private Practice	6	26.1
Years in Practice		
5 years or less	12	52%
Neuro Case workload		
70% and more	18	78.3
Neurologic Clinical Specialist (NCS)	11	47.8

Note. PT = Physical Therapist. **Participants were allowed to choose more than one option. **Participants who worked only at one of the two settings.*

Table 3.2 Descriptives of Patient Participants

Measure	*n=14*	*%*
Demographics		
Male	6	43
Female	8	57
Age (M, SD)	62.72	17.43
Time since Diagnosis (M, SD)	13.14	9.31
Use of Assistive Device	5	36
Health Condition		
PD	5	36
MS	3	21
CVA	2	14
Mix	4	29
N-MSD Dx		
1.Movement Pattern Coordination Deficit (MPCD)	4	29
2.Force Production Deficit (FPD)	5	36
3.Fractionated Movement Deficit (FMD)	3	20
4.Postural Vertical Deficit (PVD)	0	0
5.Sensory Selection & Weighting Deficit (SSWD)	1	7
6.Sensory Detection Deficit (SDD)	0	0
7.Dysmetria (DM)	1	7
8.Hypokinesia (HK)	0	0

Note. N-MSD = Neurologic Movement System Diagnosis, CVA = Cerebrovascular Accident, PD= Parkinson's Disease, MS= Multiple Sclerosis.

Table 3.3 Descriptives of MSD knowledge of Physical Therapist participants

Measure	Total	
	n = 23	%
*MSD knowledge**		
Not at all	2	8.7
I have heard about the concept, but I don't know about any particular system	3	13
I am familiar with the concept, and have heard about couple of systems or diagnostic labels, but never used any in practice	5	21.7
I am familiar with the concept and examples, and have used some neuro systems/labels in my practice, occasionally as they fit	7	30.4
I am familiar with the concept and examples and use neuro systems/labels often in practice	6	26.1
*Introduction to MSD**		
I learned about the concept decades ago when it was first introduced (~1990's)	1	4.3
I learned about specific orthopedic systems in 2000's	1	4.3
I learned about the concept within the past decade	8	34.8
I have followed the publications and correspondences on MSD since 2015	7	30.4
I have read the 2017 APTA White Paper: "Movement System Diagnoses in Neurologic Physical Therapy"	15	65.2
I have attended presentations and talks at conferences, educational settings, at work, etc.	12	52.2
I have taken courses/attended workshops on a specific orthopedic system	2	8.7
I have taken courses/attended workshops on a specific neurological system	6	26.1
*N-MSD Knowledge**		
Are you familiar with this system?	15	65.2
I have attended presentations and talks at conferences, educational settings, at work, etc.	8	34.8
I have taken courses/attended workshops on this system	2	8.7
This system was taught in my DPT/Residency program	3	13
I learned about this system when the Academy of Neurologic Physical Therapy (ANPT) developed the Movement System Task Force in 2015	12	52.2
I have read the papers published by the Movement System Task Force in 2021	12	52.2
*Use of N-MSD in clinical practice**		
I don't currently use it	3	13
I use this system occasionally in my clinical practice	2	8.7
I use this system routinely in my clinical practice	7	30.4
I am teaching this content in a DPT/residency program	8	34.8

I am involved with research in this area	2	8.7

*Note. PT = Physical Therapist. *Participants were allowed to choose more than one option.*
***Some sections included PTs that did not select any of the listed items.*

Agreement in N-MSD label assignment:

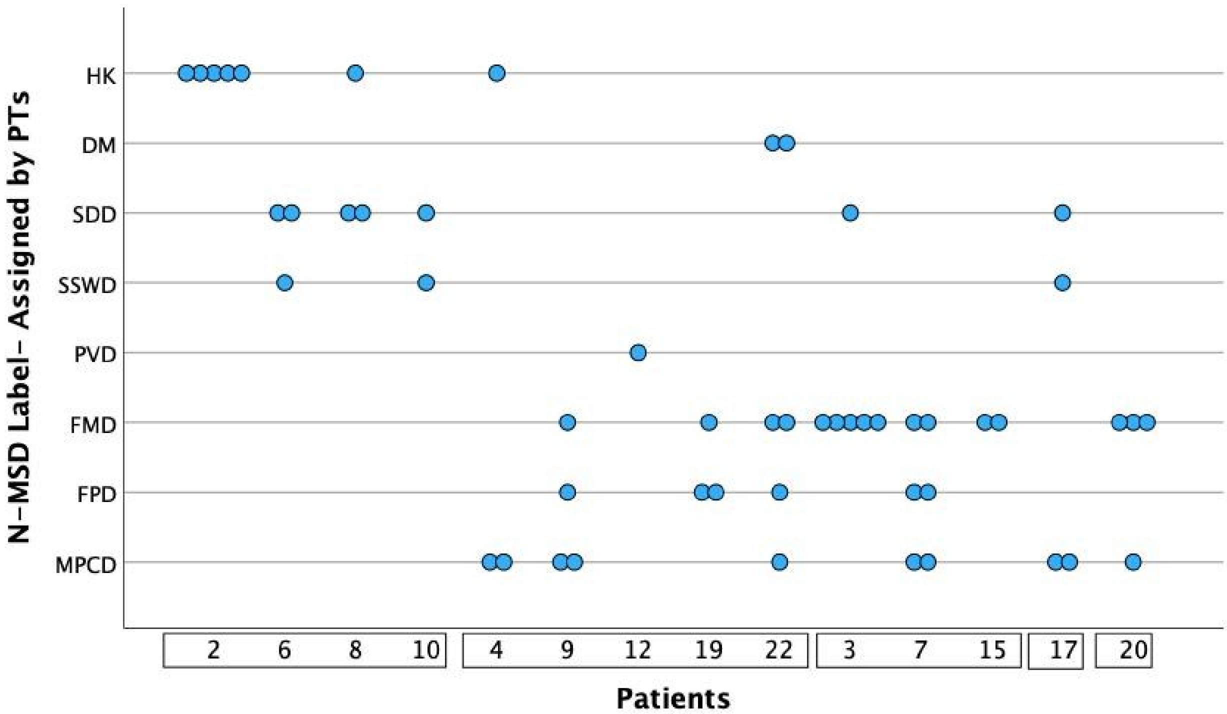
The proportion, percentage and the most agreed upon N-MSD label for each patient is listed in Table 3.4. Three patients were seen by less than three PTs; therefore, the agreement range wasn't reported for them. The range of agreement on assessments of 11 patients was between 33% and 100%. Next, all assessments of patients with the same health condition were bundled together. The weighted mean of the agreement percentages was calculated with factoring in the number of assessments in each group. The highest agreement rate was when PTs assessed patients with PD (82%). On the lower end, PTs moderately agreed when labeling patients with CVA (66.5%) with N-MSD labels. When assessments were grouped by primary movement impairments, the weighted agreement mean of 11 assessments on four patients with MPCD was 82%. The three patients with FMD impairment received assessments with variability in range, however, the overall weighted mean of agreement remained relatively high for this group (71%). The patients with FPD as their primary impairment showed 69% weighted agreement mean among PTs, across assessments and cases. The other two movement impairment groups, SSWD and DM included only one patient each, therefore no range or mean was reported for the label assignment portion of the analysis for these two groups.

Table 3.4 Agreement on Physical Therapist participants in Assigning N-MSD Labels

Agreement Among PTs	Proportion	Percentage	Most agreed N-MSD Label
I. Agreement per each patient			
Patient 2	5/5	100%	Hypokinesia
Patient 3	6/6	100%	FMD
Patient 4	2/3	66.67%	MPCD
Patient 6	2/3	66.67%	SSD
Patient 7	2/6	33.33%	MPCD, FPD, FMD
Patient 8	2/3	66.67%	SSD
Patient 9	3/4	75%	MPCD
Patient 10	2 assessments only		SSWD, SSD
Patient 12	1 assessment only		PVD
Patient 15	2 assessments only		FMD
Patient 17	2/4	50%	MPCD, SSD
Patient 19	3/3	100%	FPD
Patient 20	3/4	75%	FMD
Patient 22	3/6	50%	FMD
II. Agreement per Health Condition	**Range**		**Weighted Mean Percentage**
CVA (patients=2) [3, 7]	33%- 100%		66.5%
Mix (patients=4) [6, 9, 20, 22]	50%-75%		64.71%
MS (patients=3) [15, 17, 19]	50%-100%		71.43%
PD (patients=5) [2, 4, 8, 10, 12]	66.67- 100%		81.82%
III. Agreement per N-MSD Diagnosis			
MPCD (patients=4) [2, 6, 8, 10]	66.67%-100%		81.82%
FPD (patients=5) [4, 9, 12, 19, 22]	50%- 100%		68.75%
FMD (patients=3) [3, 7, 15]	33%-100%		71.43%
SSWD (patients= 1) [17]	-		-
DM (patients=1) [20]	-		-

Note. CVA: Cerebrovascular Accident, PD: Parkinson's Disease, MS: Multiple Sclerosis
MPCD: Movement Pattern Coordination Deficit, FPD: Force Production Deficit, FMD: Fractionated Movement Deficit, SSWD: Sensory Selection and Weighting Deficit, DM: Dysmetria.

The label assignment spread is depicted in Graph 3.1. Each dot represents an assessment that was done by PTs. The X-axis displays the patients, while the Y-axis represents the N-MSD labels that were selected by PTs. Stacked dots illustrate instances when more than one PT agreed on a label for a patient.



Graph 3.1 N-MSD Label assignment by physical therapists

1.MPCD: Movement Pattern Coordination Deficit, 2.FPD: Force Production Deficit, 3.FMD: Fractionated Movement Deficit, 4.PVD: Postural Vertical Deficit, 5.SSWD: Sensory Selection and Weighting Deficit, 6.SDD: Sensory Detection Deficit, 7.D: Dysmetria, 8.H: Hypokinesia.

Agreement in N-MSD task assessment:

To examine the degree to which ratings were identical on task assessments, the task summary scores were used. As mentioned before, the task summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. Mean, median, mode, standard deviation and range of the summary scores were reported in Table 3.5. Besides the overall agreements, assessments on patients with a similar health condition and movement impairment were clustered together for further analysis.

Table 3.5 Task Summary Scores

	Quiet Sitting						Sit to Stand						Quiet Standing					
	Min	Max	Mean	Median	Mode	SD	Min	Max	Mean	Median	Mode	SD	Min	Max	Mean	Median	Mode	SD
Agreement per Medical Condition																		
CVA (patients=2, obs=12) [3, 7]	0	3	1.17	1	1	0.94	1	11	4.67	4	3	2.74	0	3	1.25	1	1a	0.97
Mix (patients=4, obs=17) [6, 9, 20, 22]	0	1	0.24	0	0	0.44	0	9	2.88	3	0a	2.34	0	3	1.29	1	1	0.99
MS (patients=3, obs=9) [15, 17, 19]	0	5	1.22	1	1	1.56	0	14	5.44	3	1	5.41	0	6	2.78	3	0	2.59
PD (patients=5, obs=14)) [2, 4, 8, 10, 12]	0	2	0.21	0	0	0.58	0	8	2.07	1	1	2.43	0	4	0.93	1	0a	1.21
Agreement per N-MSD Diagnosis																		
MPCD (patients=4, obs=13) [2, 6, 8, 10]	0	1	0.15	0	0	0.38	0	9	1.77	1	1	2.49	0	3	0.77	1	1	0.83
FPD (patients=5, obs=17) [4, 9, 12, 19, 22]	0	5	0.65	0	0	1.27	0	12	3.65	3	3a	3.00	0	6	1.71	1	1	1.72
FMD (patients=3, obs=14) [3, 7, 15]	0	3	1.21	1	1	0.89	1	14	5.79	4.5	3	3.85	0	6	1.86	1.5	1a	1.79
SSWD (patients= 1, obs=4) [17]	0	1	0.25	0	0	0.50	0	2	1	1	1	0.82	0	1	0.25	0	0	0.50
DM (patients=1, obs=4) [20]	0	1	0.25	0	0	0.50	0	5	3.25	4	5	2.36	1	3	2.25	2.5	3	0.96
Overall Agreement	0	5	0.62	0	0	0.97	0	14	3.52	3	1	3.33	0	6	1.44	1	1	1.53

Note. SD= Standard Deviation, a: multiple values existed for the estimate. The smallest value was shown. Obs:observations.

	Quiet Standing-Feet together						Step Up						Gait					
	Min	Max	Mean	Median	Mode	SD	Min	Max	Mean	Median	Mode	SD	Min	Max	Mean	Median	Mode	SD
Agreement per Medical Condition																		
CVA (patients=2, obs=12) [3, 7]	0	4	2	2	1a	1.48	4	12	7.25	7	7a	1.96	2	11	7.25	7.5	7a	2.26
Mix (patients=4, obs=17) [6, 9, 20, 22]	0	5	1.82	1	0a	1.88	0	13	6.06	6	6	3.53	2	8	5.29	6	6	1.86
MS (patients=3, obs=9) [15, 17, 19]	0	3	0.44	0	0	1.01	0	9	3.67	4	0	3.87	0	11	4.78	6	0	4.71
PD (patients=5, obs=14)) [2, 4, 8, 10, 12]	0	5	1.43	1	0a	1.70	0	10	2.14	0	0	3.30	0	5	1.71	1	0a	1.68
Agreement per N-MSD Diagnosis																		
MPCD (patients=4, obs=13) [2, 6, 8, 10]	0	4	1.15	1	0a	1.41	0	7	1.31	0	0	2.25	0	6	1.69	1	0	1.80

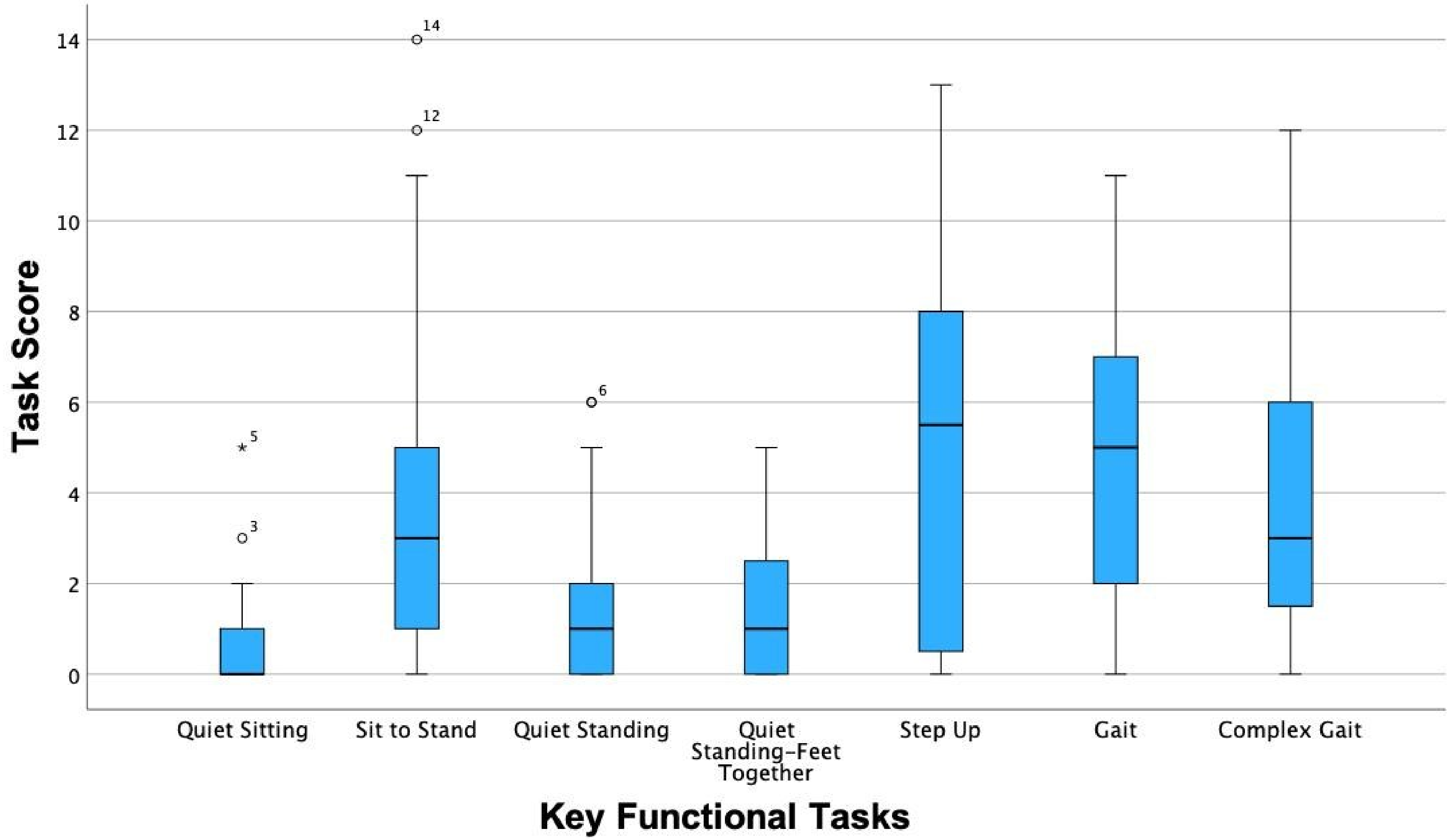
FPD (patients=5, obs=17) [4, 9, 12, 19, 22]	0	5	1.12	1	0	1.54	0	13	5.76	6	6	3.33	1	11	5.47	6	6	2.55
FMD (patients=3, obs=14) [3, 7, 15]	0	4	1.71	1.5	0	1.54	4	12	7.43	7.5	8	1.87	2	11	7.43	8	8	2.14
SSWD (patients= 1, obs=4) [17]	0	3	1	0.5	0	1.41	0	0	0	0	0	0.00	0	0	0	0	0	0.00
DM (patients=1, obs=4) [20]	2	5	4.25	5	5	1.50	6	10	8.5	9	10	1.92	5	7	6.25	6.5	7	0.96
Overall Agreement	0	5	1.52	1	0	1.66	0	13	4.87	5.5	0	3.74	0	11	4.69	5	0	3.24

	Complex Gait					
	Min	Max	Mean	Median	Mode	SD
Agreement per Medical Condition						
CVA (patients=2, obs=12) [3, 7]	1	6	3.33	3.5	4	1.67
Mix (patients=4, obs=17) [6, 9, 20, 22]	1	12	4.53	4	2	3.34
MS (patients=3, obs=9) [15, 17, 19]	0	3	0.78	0	0	1.09
PD (patients=5, obs=14)) [2, 4, 8, 10, 12]	1	9	5.14	5.5	1a	2.57
Agreement per N-MSD Diagnosis						
MPCD (patients=4, obs=13) [2, 6, 8, 10]	1	12	4.77	4	4	2.98
FPD (patients=5, obs=17) [4, 9, 12, 19, 22]	0	11	3.88	2	2	3.64
FMD (patients=3, obs=14) [3, 7, 15]	0	6	2.86	3	4	1.96
SSWD (patients= 1, obs=4) [17]	1	3	1.75	1.5	1	0.96
DM (patients=1, obs=4) [20]	4	6	5.25	5.5	6	0.96
Overall Agreement	0	12	3.77	3	4	2.88

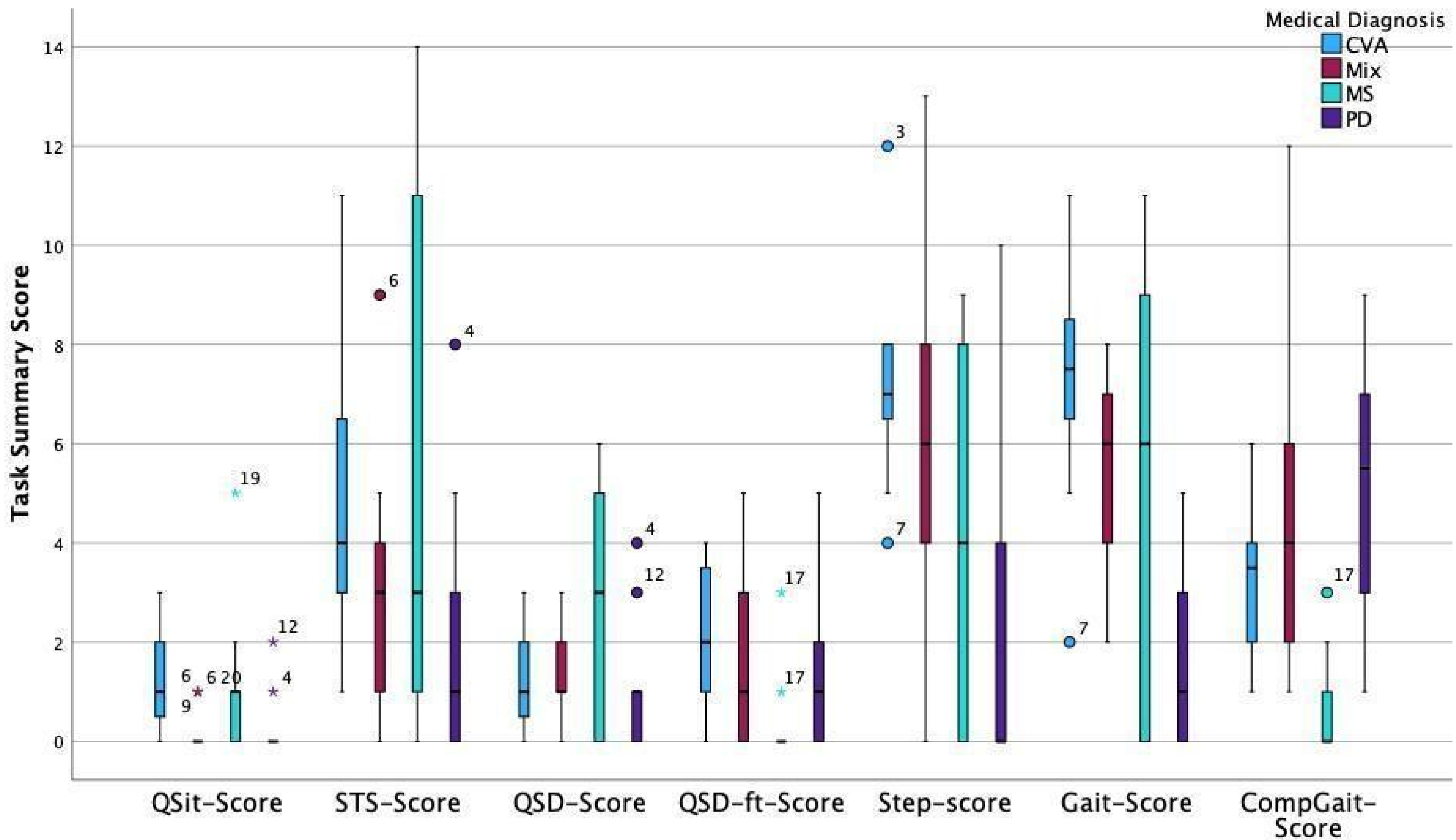
The following graphs show the task summary score analysis across all patients (Graph 3.2), per health conditions (Graph 3.3), and movement impairments (Graph 3.4). The bold line in the box represents the median, while the lower and upper outline of the box show the low and high quartile of the summary task score. If outlier or extreme values were given, they were shown by dots outside of the box range.

The overall task summary scores across all patients, displayed in Graph 3.2 as well, showed that the mean of summary scores was lowest for static tasks, Quiet sitting (mean= 0.62, SD= 0.97), followed by Quiet Standing, and Quiet Standing with feet together tasks. The highest mean and median of task scores belonged to Step Up (4.87, 5.5), indicating that PTs were identifying more deficits while watching patients' step up performance. However, the scores on the Step Up task were dispersed with a wide range (0-13) and highest standard deviation (SD= 3.74) indicating variability in ratings among PTs. In contrast, Sit to Stand and Complex Gait tasks were also marked often with existent deficits, but did not show the same amount of variability as displayed during the Step Up task (SD=3.33; 2.88 respectively).

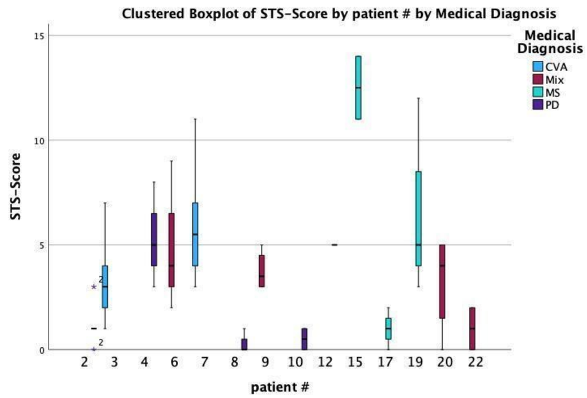
In further breakdown of task scores based on health condition (Graph 3.3), the ratings on patients with multiple sclerosis (MS) (3 patients, 9 assessments) and mix neurologic diagnosis (4 patients, 17 assessments) showed the highest amount of variability across most of the tasks. For patients with MS, specifically, the level of agreement among PTs on the tasks of Sit to Stand, Step Up and gait was very low. Patient by patient analysis showed that the high variance in the scoring of the three patients with MS was caused by between-patient difference and not within-patient rating differences. In all three tasks (Graph 3.4 (a), (b) and (c)), patient 17 was marked with almost no deficit, while the other two patients with MS, 15 and 19, were marked on 5-10 deficit items on the assessment forms. For patients in the mixed neurologic conditions group, the range of scores as well as standard deviation were highest for Step Up (0-13, SD= 3.53) and Complex Gait (1-12, SD=3.34) tasks. Patient by patient analysis showed that the variability in ratings was caused mostly by between-patient differences. On Graph 3.4 (b) and (d), patient 9 ratings showed a wide range of deficit marks by PTs, while the other three patients (13 assessments) showed a less than 5 points difference across most tasks.



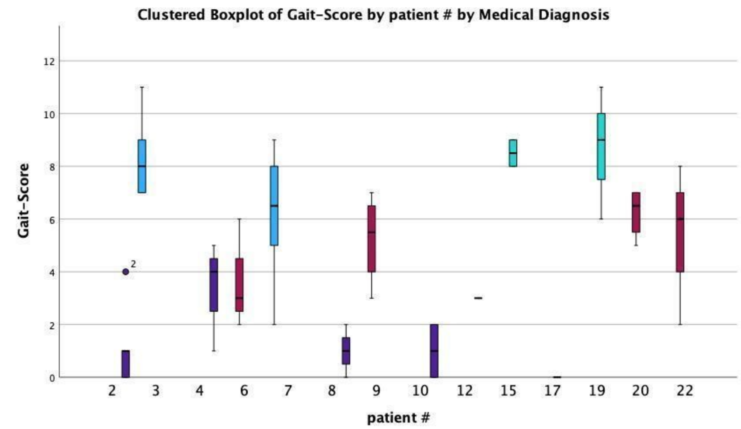
Graph 3.2 Task Summary Score



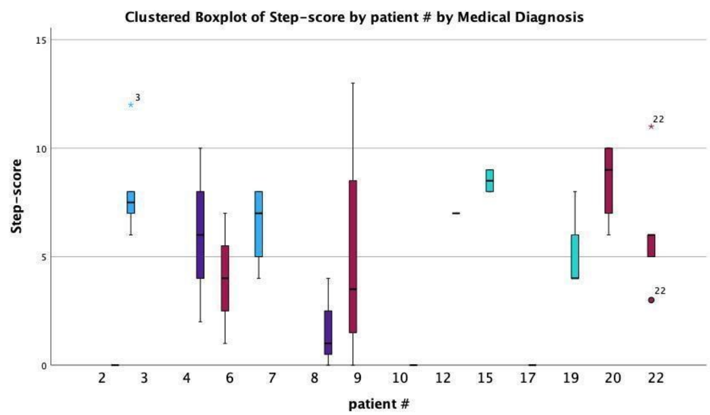
Graph 3.3 Task Summary Scores per Health Conditions



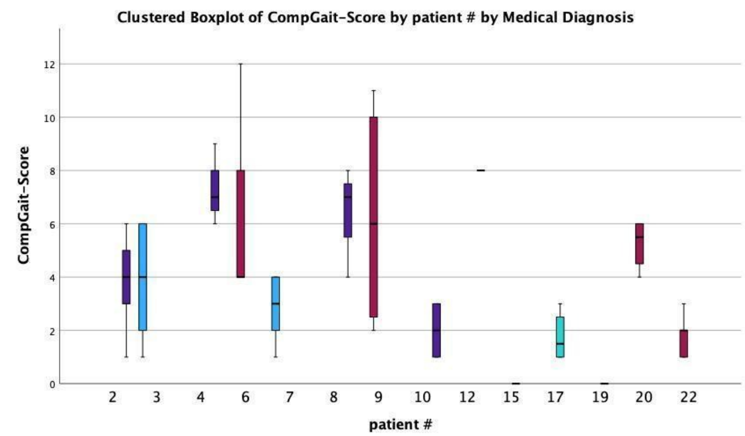
Graph 3.4 (a). Sit to Stand



Graph 3.4 (c). Gait

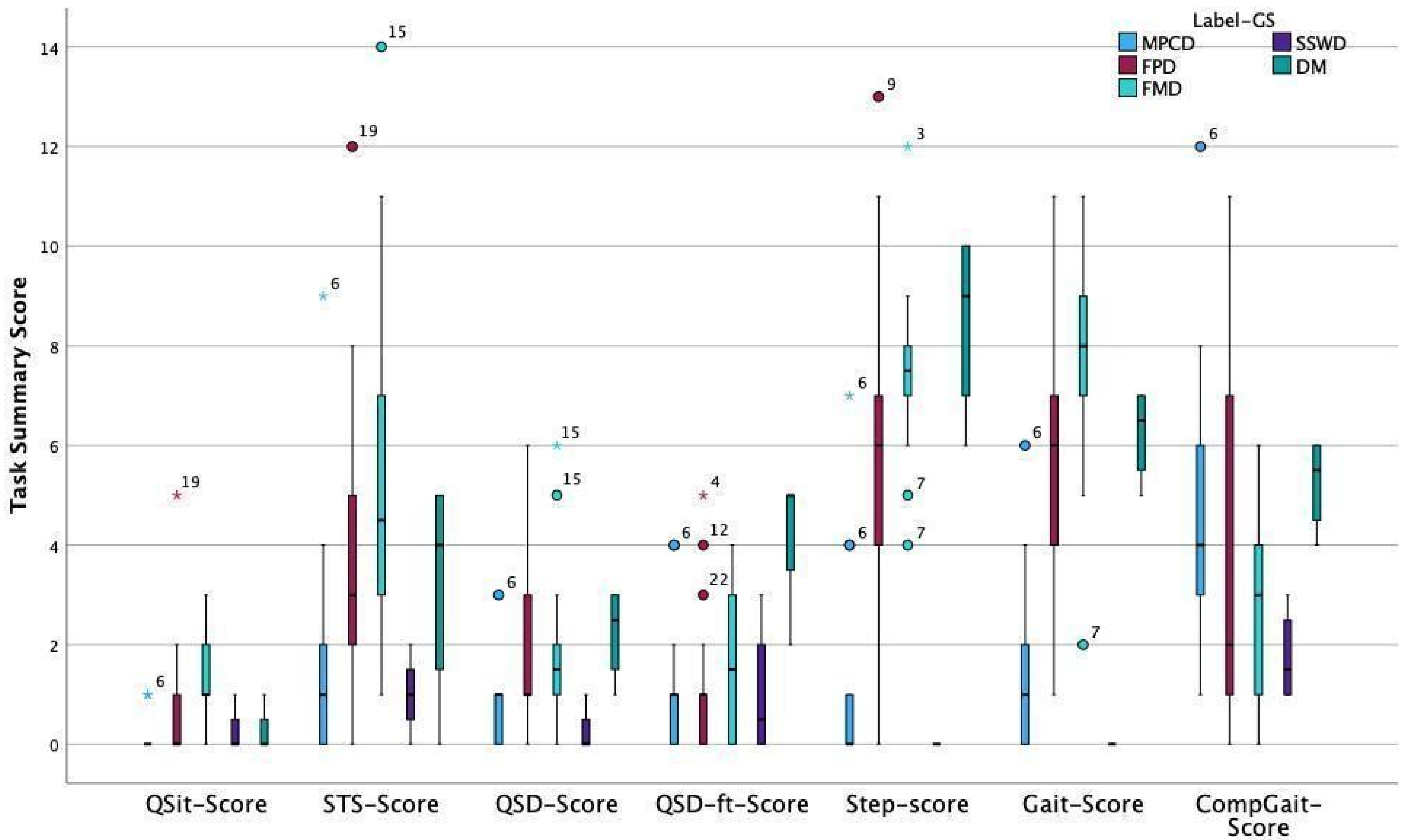


Graph 3.4 (b). Step Up

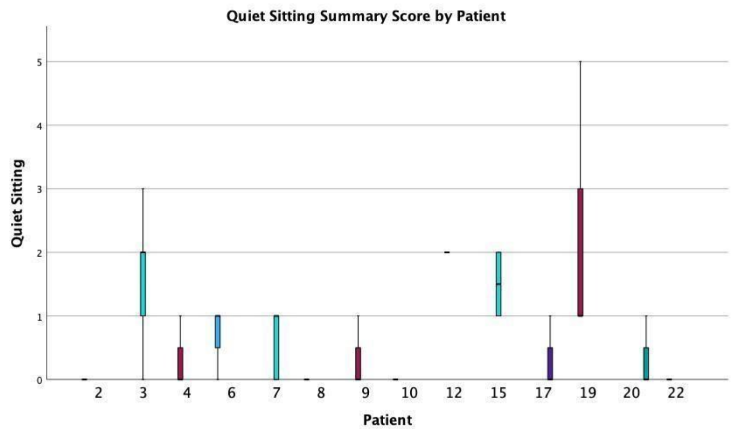


Graph 3.4 (d). Complex Gait

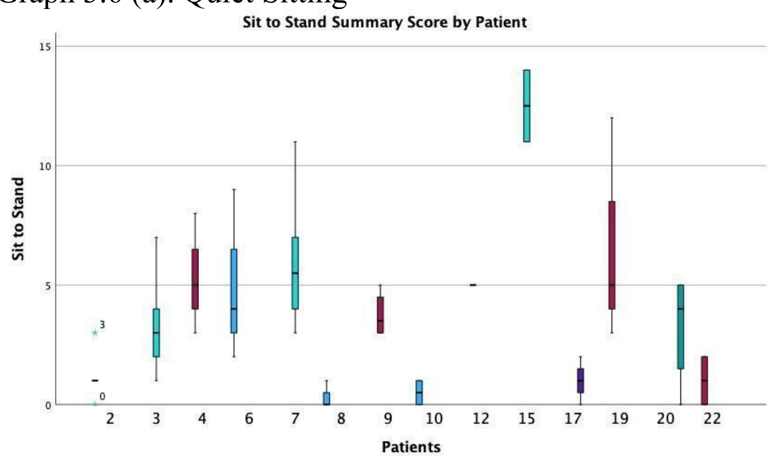
In further breakdown of task scores based on primary movement impairments (Graph 3.5), the variability in ratings of patients with FPD impairment (5 patients, 17 assessments) was consistently high across tasks (depicted in red) compared to other groups. Patients with FMD impairment (3 patients, 14 assessments, depicted in turquoise) were marked with a wide range of scores during Sit to Stand (1-14, SD=3.85), and several outlier scores during Quiet Standing, Step Up and Gait. For patients with FPD, the source of the score variability is different for each task. During Sit to Stand, patient 19 was scored with minimum of 3 and maximum of 12 deficits (Graph 3.6 (b)). Patient 9 was marked with 0 to 13 deficits for Step Up task (Graph 3.6 (e)) and 2 to 11 deficits during Complex Gait task (Graph 3.6 (f)). The dispersion of scores on gait were due to between-patient differences within the FPD group. For patients with FMD, the source of score variability during Sit to Stand and Quiet Standing (Graph 3.6 (b) and (c)) was between-patient differences, as patient 15 (MS) was scored with a significantly higher number of deficits during these two tasks. The same pattern was observed for Step Up task, as one patient was scored with only a couple of deficits while another patient received higher points on the assessment form.



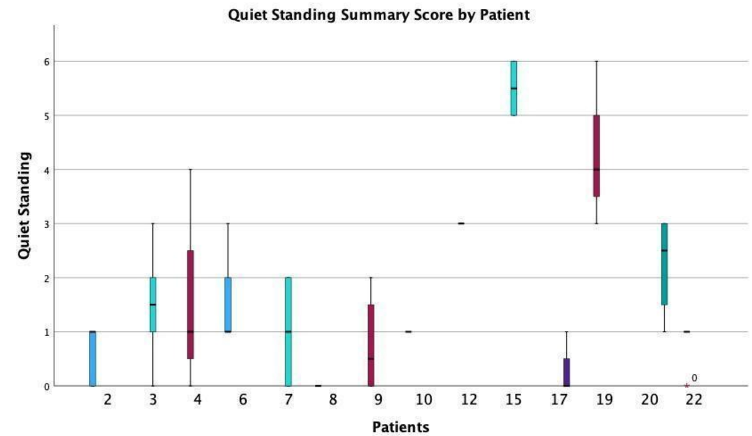
Graph 3.5 Task Summary Score per Movement Impairments



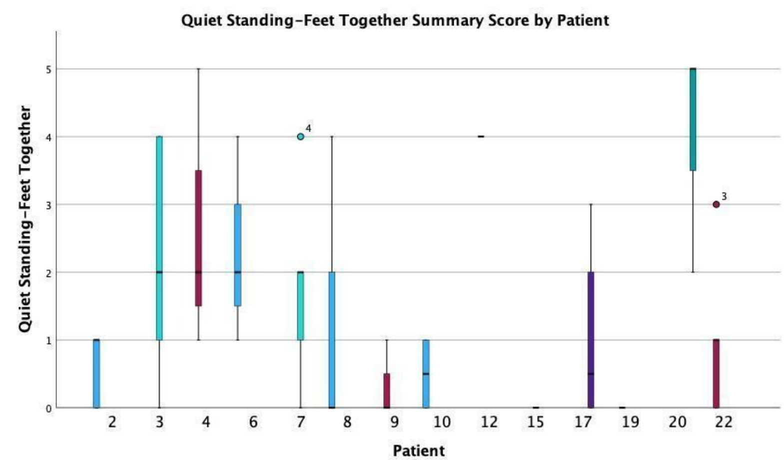
Graph 3.6 (a). Quiet Sitting



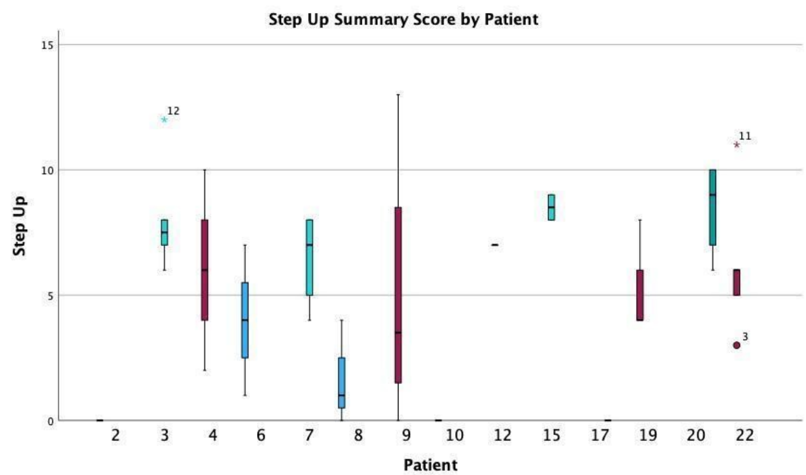
Graph 3.6 (b). Sit to Stand



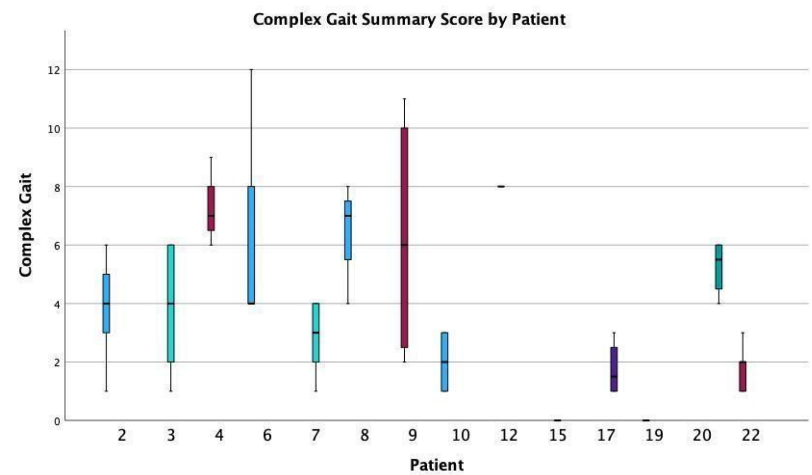
Graph 3.6 (c). Quiet Standing



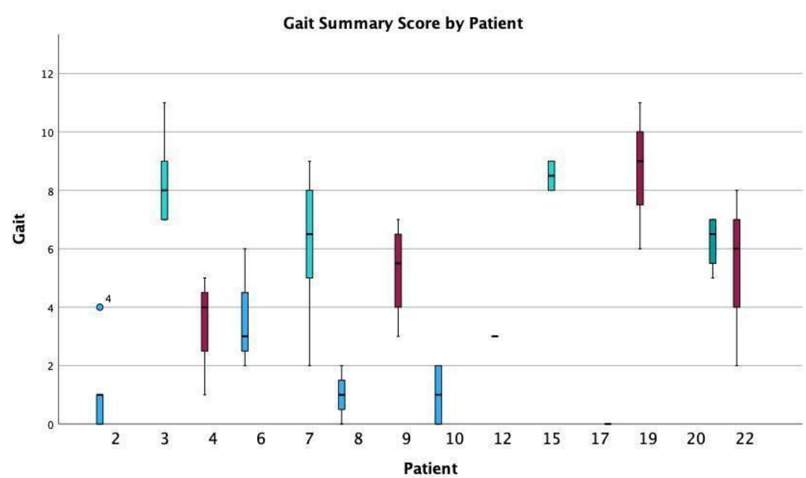
Graph 3.6 (d). Quiet Standing-Feet Together



Graph 3.6 (e). Step Up



Graph 3.6 (g). Complex Gait



Graph 3.6 (f). Gait

3.4 Discussion

A prerequisite for using a Movement System classification framework is for therapists to be able to reliably examine functional skills and group patients into the suggested movement system diagnostic categories. The purpose of this study was to examine the agreement among PTs in using the N-MSD framework in patients with neurologic health conditions.

The label assignment analysis showed that therapists with different levels of clinical experience and previous familiarity with Movement System frameworks could assign the same diagnosis to patients with the same primary movement deficits, with agreement rate of 69% to 82%. However, when the same patients were assessed by an expert, often a different N-MSD label was assigned to the group of patients. For example, patients that were identified with MPCD by an expert received Hypokinesia, SSD, and SSWD labels by the PT participants. This difference could be partially explained by how therapists usually associate a diagnosis with a certain medical condition. For example, patients with PD were often labeled with Hypokinesia as the main movement issue, while the same patient was identified by MPCD or FPD by the expert.

Aside from the label assignment analysis, agreement in task assessments was also examined in this study. Analysis of the overall task scores showed that the stationary tasks such as Quiet Sitting and Standing consistently were marked with fewer deficits present compared to transitional and dynamic tasks such as Sit to Stand, Step Up, and Gait. This finding was well explained by the fact that by design, there were fewer items listed for Quiet Sitting (6 deficits), Quiet Standing (9 deficits) to begin with. Moreover, the relative smaller variability in rating of these two tasks and therefore higher agreement compared to other tasks were partially due to the included participants. The average age of patient participants was 63 years, with 14 years since diagnosis on average. Besides the long-term condition, most of the patients had high level of functionality, with only five using assistive devices. Therefore, lack of functional deficits during stationary tasks was an expected finding.

Analysis of the agreement among PTs in assessment of transitional and dynamic tasks did not show a consistent pattern, as agreement was high for some groups and low for others. In two occasions, four PTs assessed the only patient falling under a primary movement impairment (one patient with DM and one patient with SSWD labels). Given that there was no between-patient variance among the ratings in either of the groups, the within-patient rating variability was purely due to differences in PT's scoring of the patient. The PTs showed a relatively high level of agreement in scoring of all the tasks (SD < 1.00), with exception of Sit to Stand and Step Up tasks. The agreement on the final N-MSD label was moderate in these cases (patient 17, four assessments: 50% agreement rate; patient 20, four assessments: 75% agreement rate). These patients were different in both level of function and severity of their conditions, as one was physically active in the community with minimum endurance impairments while the other one was limited in community walking heavily relying on another person close by (Patient 17: MS, walked independently with no assistive device. Patient 20: Mix- traumatic brain injury (TBI), no assistive device but relied on using furniture indoor and a standby level of assistance from another person when walking outdoors. Coordination was heavily impacted during complex dynamic tasks)

The noticeable amount of variability in ratings of patients within a certain movement impairment group or a health condition often came down to two to three “hard to rate” patients. For example, FPD and mix neurologic conditions groups showed the second most difference and dispersion of the ratings. However, patient-by-patient analysis showed that a single patient (patient 9) was the source for a significant amount of disagreement among PTs. One explanation for the discrepancy of ratings of this patient is that the neurologic condition of amyotrophic lateral sclerosis (ALS) impacted the function of both sides. When assessing Step Up and Gait tasks, some PTs could have missed some deficits as they were considering one side as the more affected. It is possible that patients with a bilateral presentation in movement deficits, regardless of condition, may be more challenging to observe, may have more subtle findings, and therefore, may be more difficult to diagnose.

Another source of variations in some tasks, especially complex gait tasks, could be explained by what was considered as performance of an item. In patient videos, minutes of a few trials of a task or a group of activities related to a task group were bundled together and labeled with the corresponding task. However, each item related to a task might have not been labeled individually. For example, stepping over obstacles, backward stepping, stepping in various sensory conditions, etc, are all items on the complex gait task. However, because the patient's performance might have not been identified either verbally by the examiner or via caption, some therapists could have missed individual items related to a task. Therefore, a task score, identified by the number of the deficits selected on the assessment form, could be very different between PTs who identified the various items in the videos and those who did not.

Variability in observational ratings of core tasks in this study were in line with some evidence found in the literature. A systematic review on observational gait assessment measures showed that while structured forms improve consistency, their effectiveness is limited due to non-standardized protocols and subjective interpretation (Ferrarello et al., 2013). With lack of the precision needed for objective evaluation, the need for validated, standardized tools and methods were highlighted. Adopting the Movement Continuum format and including detailed checklists in N-MSD observational task assessment forms provided a comprehensive yet uniform measure for core task evaluation. However, based on the variability in the assessments, a focused training on these task checklists seems to be an imperative issue prior to testing reliability. Thus, a more rigorous examination by the PT as well as proper training on task assessment elements could enhance the clinical utility of the N-MSD task assessment and consequently a movement label assignment.

Limitations

This study had some limitations. The N-MSD framework was designed for PTs to do a thorough in-person physical examination of patients before making diagnostic decisions. In this study, PT participants were required to make assessments based on written impairment findings and short videos of patients performing functional tasks. While the GS for examining movement dysfunction is 3-dimensional (3-D) movement analysis, use of 2 dimensional (2-D) videos of patient performance that could be also replicated in clinical settings have been established as a reliable and validated method. Several studies provided evidence for moderate to high reliability for observational gait assessment through use of videotapes (Brunnekreef et al., 2005; Ferrarello et al., 2013). A study showed that by watching visual 2- D videos of athletes performing single leg squat, novice and experienced PTs were able to classify lower extremity movement patterns in patients with musculoskeletal disorders (Harris-Hayes et al., 2014).

The examination process and commands used in the patient functional task assessment were not standardized in our study. The examiner relied on the N-MSD manual for required adjustments, facilitation, and commands during tasks performances, however, there were inconsistencies in execution of the tasks. Partially caused by space, time, and patient ability constraints, the examiner's discretion about modifications to the procedures may have introduced variations in task performances. Moreover, when reviewed by the GS, several assessments needed more in-depth investigation of findings that the examiner could have followed up with further testing.

The patient participants in this study were recruited from outpatient physical therapy clinics in the community. Hence, most of the participants were community-dwelling older adults capable of ambulating independently or with or without use of assistive devices or some assistance. This made the study sample limited in diversity in terms of individuals' function level, with no severe or acute case representation.

Gold standard reliability measures such as Kappa and Interclass Correlation Coefficients (ICCs) were not measured in this study due to the small sample size. Specific statistical approaches for assessment of reliability and agreement have been developed within different disciplines. Thus, there is no single statistical method to consider widely accepted to follow. It is also encouraged to report a combination of estimates when reporting the results to represent different statistical methods (Kottner et al., 2011). Nevertheless, the agreement proportions, percentages and range, although easily understandable, are limited to be extended for reliability purposes, as they do not adjust for agreements by chance. Other forms of reliability measures such as test-retest reliability are also recommended to be studied in future studies.

3.5 Conclusions

The shift towards use of MS frameworks and diagnoses in physical therapy is becoming more prevalent. This study examined the agreement among therapists in using an example of these frameworks, N-MSD. Our findings showed that PTs agreement on assigning N-MSD labels to patients with PD and MPCD as their primary movement impairment was the highest compared to other groups.

Observational movement analysis is also an integral part of the N-MSD frameworks. The examination of the structured forms, designed for assessment of movement deficits across stages of a task, showed a high level of agreement among PTs in assessment of static tasks such as Quiet Sitting and Standing. More variability in scoring of dynamic and complex tasks were shown overall. However, transitional tasks Sit to Stand and Step Up displayed the highest levels of variability in PTs' assessment compared to Gait and Complex Gait.

To integrate N-MSD into clinical practice, future studies should investigate other psychometric properties of this framework across patients with various health conditions. Adequate sample size and more unified training on the framework could enhance the robustness and applicability of the results.

.

Chapter 4. Inter-rater agreement between experienced and early-career physical therapists in using Neurologic Movement System Diagnosis (N-MSD)

4.1 Introduction

PTs assess and treat patients with movement impairments on a daily basis. In the physical therapy literature, classification systems were developed for various purposes such as education, research, and clinical practice within different physical therapy care settings. In the early decades of the profession, the traditional pathoanatomical medical frameworks were widely used, anchoring on the available science. However, by the turn of the century, leaders advocated for more evidence-based approaches to be used in educational and practical settings (Deutsch et al., 2022; L. P. Scheets et al., 2021).

In 2013, The American Physical Therapy Association (APTA) adopted movement optimization as the core of the physical therapy practice (*Vision Statement for the Physical Therapy Profession and Guiding Principles to Achieve the Vision*, 2013). To optimize movement, it is necessary to systematically identify the root of movement dysfunctions and target the underlying causes rather than signs and symptoms (*Vision Statement for the Physical Therapy Profession*, 2019). While the idea of movement-specific system labels is not new to our profession, there is a lack of rigorous studies that examine and support the reliability, validity, and clinical application of these systems.

The N-MSD by Scheets et al. is among examples of available MS frameworks. First published in the 1990's, this system was developed through years of working with various patients with neurologic conditions and therefore could be used with more than one specific patient population. Since the introduction of this system, the developers published case studies and research articles to demonstrate the application of this classification system in a variety of patient populations with movement dysfunction resulting from neurologic health conditions (Scheets et al., 2015).

The N-MSD, an example of a structured and well developed framework for MS approach, consists of key tests for impairment in body function as well as observational analysis of critical functional tasks that are performed by patients. The clinical tests used during the physical examination part of the N-MSD protocol are widely used in standard physical therapy examination sessions. These tests were designed to identify deficits in motor control, muscle tone (level of hyperexcitability), muscle strength (force generating capacity), nonequilibrium coordination, sensation, postural control, motion sensitivity, mental status, and joint range of motion. The psychometric properties of these clinical tests were previously established within the physical therapy literature. The physical examination follows by assessment of key functional tasks including Quiet Sitting, Sit to Stand, Quiet Standing with feet hip's width apart and with feet together, Step Up (i.e., placing one foot on a step and returning it to the floor), Walking, Walking while turning the head, stepping over obstacles, and walking forward and backward. These core functional tests were selected because they measure movement variables that are key to differentiate among MS problems. In an attempt to standardize the assessment of core functional tasks, the N-MSD adopted the Movement Continuum format for observational movement assessment (Hedman et al., 1996). The tasks were broken down into stages of movement: initiation, execution, and termination. The diagnostic element of the function-related tests is the observation of how a person accomplishes the activity, not the level of physical or verbal assistance needed. A checklist of possible movement deficits related to sequencing and timing of movement of body segments, body weight support, postural stability, and limb/trunk alignment are provided for various phases among each core task. During observational analysis of core tasks, PTs select the deficits (from this checklist) present with the patient.

Together, the examination components enable the PTs to differentiate among movement system problems. The results from these tests and observational task analysis guide PTs in identifying the primary movement dysfunction. Aligned with the advantages of physical therapy-specific classification systems discussed before, the diagnostic labels in this system groups patients based on similarity in movement impairments rather than medical pathology or health condition.

Experience is crucial for developing more sophisticated skills in both movement assessment and clinical decision making (Jensen et al., 1990, 2000). Experienced PTs demonstrate more integrated and forward-thinking approaches, while novices tend to be more linear and reliant on protocols and external information (Wainwright et al., 2010, 2011). With observational movement analysis in particular, expert PTs tend to adopt a more fluid and integrated approach compared to novices (Chua et al., 2024). Experienced PTs employ a holistic and integrated approach, observing movement at multiple levels and cyclical manner. Their analysis involves continuous assessment, treatment planning, and re-evaluation to synthesize information. In contrast, novices rely more on analytical approaches, systematically evaluating each component of movement due to limited practical experience.

Similarly, there are differences between the two groups in terms of Clinical Decision Making and selecting diagnosis. Experienced PTs have a broad and structured knowledge base, utilizing forward reasoning and pattern recognition to develop a working diagnosis. On the other hand, novice PTs often rely more on hypothetico-deductive reasoning and external resources to guide their decisions. They have a more limited ability to integrate information and anticipate outcomes due to less experience and conceptual knowledge (Wainwright et al., 2010, 2011).

Considering the recommendations suggested by the ANPT MS Taskforce, the first step in the widespread implementation of MS diagnoses is to study the psychometric properties of extant systems, especially in the context of the clinical experience of the examiner. Additionally, while the face validity of the N-MSD was examined during the development and repeated clinical use, no controlled research study has examined the validity and reliability of the entire system (Hedman et al., 2018) or the differences seen in ratings between experienced and early career PTs.

The purpose of this study was to examine the agreement in ratings between early-career and experienced PTs, in analyzing different tasks based on the N-MSD and assigning diagnostic labels in individuals with various neurological health conditions. We expect to see higher agreement in experienced PTs in assignment of labels and movement assessment.

4.2 Methods

The study was designed to take place in four overlapping phases. All procedures were approved by the University of Washington's Institutional Review Board. The overall goal was to establish the agreement between ratings by PTs and to those from the GS expert. The GS in this project was the primary person in the group of clinicians and researchers that developed the N-MSD. To meet the aims, the first step was to prepare and record patient videos and develop gold standard profiles for each patient. Next, recruited PTs were asked to review the patient profiles and assess and rate them using the N-MSD framework. The study roadmap is presented in Figure 4.1.

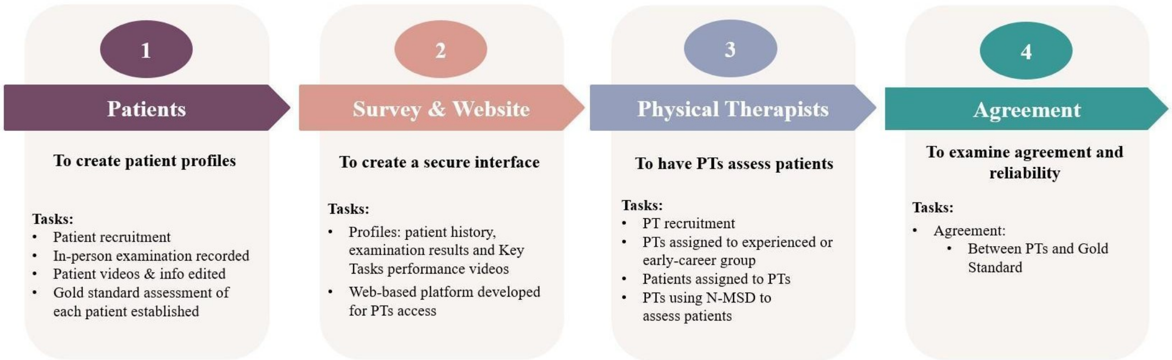


Figure 0.1 Study Roadmap

4.2.1 *Phase I: Patient Participants*

In preparation for patient recruitment, three pilot examination sessions were done to map out the camera setups and testing order. After patient screening, recruitment, and in-person examination sessions, profiles were created for the patients, reviewed, and edited for further use in the study. Details of these activities are outlined below.

Data collection forms; REDCap© platform

Data for this study were entered into and/or collected by instruments on a secure web platform, REDCap© (Research Electronic Data Capture) hosted at University of Washington (Harris et al., 2009, 2019). REDCap© is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. Surveys and instruments were designed and built in REDCap© by the Primary Investigator (PI) with assistance from administrative research organizations within the UW.

Clinic/lab environment and the study session

The setting for the in-person examination sessions was at the same physical therapy clinic that patients regularly visited. A couple of patients were seen at a research lab on the University of Washington campus. Two cameras were available to capture anterior-posterior and lateral views for all functional tasks.

Patient participants

Recruitment

The subjects were recruited from different sources such as physical therapy clinics, community centers, support groups, and through flyers placed in outpatient clinics across Oregon and Washington state. Potential subjects were contacted by email or phone calls initially. Subjects verbally agreed to be contacted prior to the initial contact by the PI. After being presented with a brief explanation about the study aims and upon their indication of interest, they were asked to fill out an online screening survey. The survey questions confirmed eligibility and asked for further demographic information. Adults with neurological conditions with the ability to walk 20 feet were eligible, regardless of level of assistance needed. Individuals who were not able to successfully ambulate 20 feet were not invited to participate in the study. The eligible individuals were contacted and scheduled for a one-time examination session.

Examination procedures

The examination sessions were conducted in a quiet environment with no interruption from other patients or researchers. The cameras were moved as needed during the sessions to have a frontal and lateral view of the patient at all times. Occasionally the room size did not allow for both views in which case the best angle was chosen by the examiner (PI).

The cameras were managed by the same examiner (PI) and were turned on at the beginning of the session after written consent was obtained. The sessions started with open-ended questions about the participant's medical history followed by routine physical exams. Patients were seated on a chair with no arm rests or at the edge of the bed with feet on the floor for most of the tests. Tests were done in a gradual manner based on the guidance provided within the N-MSD manuals, but occasional adjustments were made as needed. According to the N-MSD manual, the examiner was to provide as little assistance as possible during all the functional tasks. In cases where participants were not able to complete a task, either limited by physical ability or fear of falling, the minimal amount of assistance required was given. Simpler functional tasks such as Quiet sitting, Sit to Stand, and Quiet standing were sometimes examined during impairment testing. Patients were offered water and rest breaks during the exam and the session concluded with more complex gait tasks. Participants were given $40 in the form of a Visa gift card at the end of the session.

Creating patient profiles

The patient data was initially curated and edited by the PI as described below. The results of the body structure and function tests were entered into REDCap© intake forms. Along with patient history and other relevant information, all the documented written data was compiled into a single portable document file (pdf). Next, footage from the two cameras were overlaid and adjusted for both picture frame and audio to prepare a single video clip of all functional tasks. For more clarity, solid and titled frames separated functional tasks from each other on the clip. The audio was regularly muted on videos for several reasons, mainly to protect patients' privacy or hide the conversation between patient and the examiner. If a verbal cue was given to the patient during the muted footage, closed captions were provided on the video to summarize what guidance was given to the participant. Given that only functional task performances were included in the video clip, the final edited video was usually less than 15 minutes in length (the shortest video being 5:27, and the average video length was 8:30). Due to a camera malfunction, the Step Up task in one video was recorded from posterior view only, and gait task was only recorded for a few seconds.

A profile for each patient was created, consisting of the written data pdf file and core functional tasks video clip. The GS was given access to all participant profiles. Except for the initial profiles, the feedback was provided through asynchronized recorded sessions from the GS shared screen. On average, two profiles were reviewed in one sitting and returned with detailed feedback on the video frames of each functional task. In most cases, videos were edited to contain only relevant footage to the N-MSD framework. Suggested captions were added, and unnecessary minutes were removed. For some patient profiles, short clips of various additional tasks or tests, besides the core functional ones, were kept and labeled as extra clips. The decision was made jointly by PI and GS after reviewing each profile. While these extra minutes did not contain essential tasks from the MSD manual, there were diagnostic values to certain activities observed during that led to the decision of adding them to the patient examination videos.

Gold Standard assessment creation process

The GS completed the N-MSD assessment form while reviewing the data and chose a N-MSD diagnostic label at the end. In some cases, when a second diagnosis was applicable, it was recorded to be considered for further analysis. The result was saved as the GS assessment for each patient to be used later in the reliability examination phase of the study.

4.2.2 *Phase II: Website*

The goal of phase II was to a) develop an interactive online platform to upload and store all the patient profiles on, and b) create training modules on the N-MSD examination and diagnosis.

Website and The Patient Profile Pages

To securely share the patient data with PTs in the second phase, multiple options were considered. After considering the cost-benefit of developing an independent website, the research team decided to use a cloud-based collaborative platform for this study instead. Microsoft's SharePoint is an example of a cloud-based system that allows members to store, share, and manage data while providing various levels of security and access for each member. Additionally, creating a site through SharePoint does not require language-based coding. The site consisted of two major sections: patient profiles and training. Short videos of 2-3 minutes with explanatory narration were posted to orient users to the template of different parts of the website.

On each profile page, demographic information, history, and body function and structure test results were presented followed by the key functional tasks video and finally a link to the assessment form. There was no time limit to complete the assessment and participants could look at the patient information at any part of the assessment.

The Training Page

The training content was developed and reviewed by the GS and PI. The modules were in the format of PowerPoint© slides, embedded into the study training page. As PTs were expected to have different levels of clinical experience and MSD knowledge, the modules were introduced as a refresher and/or training material. While all PT participants were asked to complete the modules, each determined the pace and amount of time to spend on the material. We expected that on an average, participants would take 1-2 hours to complete the training, depending on their clinical experience, previous exposure to any MSD classification system, and established familiarity with N-MSD.

There were three training modules:

1. Introduction to MSD: This module covered an overview of the Movement System (MS) concept, the history and dialogue around the topic, and provided a brief list of other systems that were created and used throughout the decades in the physical therapy literature. The need for a Neurologic MS framework and early versions of N-MSD were also explored.

2. N-MSD diagnosis: A set of six slides were developed for each of the diagnoses, consisting of clinical presentation, considerations related to prognosis, treatment focus, and likely associated underlying pathophysiology. For each diagnosis, videos and/or photos of a case example were integrated. The last slide of a diagnosis had a knowledge checkpoint quiz. Questions on these knowledge checkpoints were written by the GS and intended to assess if participants were comfortable with the key elements and basic concepts relevant to the diagnosis. The checkpoints were not graded, and participants could see the correct answers immediately after completing the quiz.

3. The N-MSD assessment form: Adapted from the original N-MSD manual (published as Appendix to Scheets et al. 2014), a detailed assessment form was developed to include all the key functional tasks. Typical performance of each task was defined by the essential movement components for that task. Additionally, expected movement deficits for that task for each diagnosis was described. For transitional tasks, the observations were designated by phase of movement (initiation, execution, and termination). The number of possible movement deficits for each task varied, but items making up a task phase were separated by the phase label. In module 3 of the training page, the assessment form was available so that participants could familiarize themselves with it before beginning patient assessment.

Besides the training modules, supplemental documents of the original N-MSD publications were included on the training page. The PTs had access to the examination manual and diagnosis tables throughout the study.

4.2.3 *Phase III: Physical Therapist Participants*

The goal of phase III was to recruit PTs, have them review patient profiles, and assess patients using the N-MSD framework.

Recruitment

Individuals who worked in neurological settings and/or had 70% of their caseload focused on patients with neurologic health conditions were selected for this study.

The early-career PTs were sought through various methods such as contacting a) alumni of several Doctor of Physical Therapy (DPT) programs, b) neurologic residency programs for either current or former residents, and c) individuals suggested and referred by research team members and d) clinical instructors or other community PTs. Participants were deemed early career if they had less than five years of experience.

Experienced PT Participants

Individuals with years of experience in physical therapy practice with some knowledge of Movement Systems were sought out for this group of participants.

The experienced participants were recruited from multiple sources by the following methods: a) recommended by study team, b) through inquiry from former or current members of several ANPT task forces, c) by contacting directors of several neurologic residency programs, and d) recommended by PTs who participated and completed the study. Participants were deemed as experienced PTs if they had more than five years of clinical experience.

After the initial contact by the PI, all candidates were asked to complete a screening survey. Individuals were asked about their current working status, clinical workload and its patient population makeup, the length and history of their clinical practice, experience in neurologic settings, familiarity with the Movement Systems in general, any of the MSD frameworks and N-MSD in particular. Based on the answers provided, individuals were invited to participate in the study and assigned to either the early-career or experienced PT group.

Profile assignments to PTs

Two patients with different health conditions and N-MSD diagnoses were randomly assigned to each PT. Thus, each patient was seen by at least two PTs. In a few cases a PT reviewed more than two profiles, thus a patient was seen by more than four therapists total.

Access to the Website

Participants received website invitations upon inclusion criteria fulfillment. When accepting the invitation, they were added to the University of Washington (UW) cloud directory through a Microsoft email proxy. Assigned patient and PT participant identification (ID) numbers were emailed to each participant before website access. Participants were originally asked to complete the study within a 2-3 week period. Once the participation was underway, the PI checked on the progress periodically and sent reminders when needed. One participant only completed one out of two patient assessments.

4.2.4 *Phase IV: Examine agreement and reliability*

An estimation for sample size for both patients and PTs was suggested. This estimation was primarily based on similar previous studies of classification systems in the musculoskeletal literature

Data analysis

Data cleaning was done by the PI concurrently with the data collection. Eighteen subjects with a range of neurologic conditions were examined by the PI. After the recruitment and examination sessions, four patients were excluded from further stages of the study. Two of these individuals required more than 50% assistance (moderate to maximum assistance) for ambulation, one needed assistance from more than one person for more advanced skills such as Step Up and Gait. Another individual experienced mood and emotional disturbances during the first three functional tasks and the examination session had to end abruptly. Data from 14 participants were used in this study.

Reliability and agreement are both important issues when developing and testing new classification systems. Agreement is the degree to which measurements or ratings are identical (Kottner et al., 2011). Therefore, to examine the similarity between PT and the GS ratings in using the MSD framework, agreement estimates were considered appropriate. The type of variables involved in the analysis (categorical versus continuous) are other factors to consider when choosing estimates of agreement. To report agreement on a nominal variable, such as N-MSD labels, proportion and percentage agreement were reported. Agreement among raters on continuous measures, such as core task scores, range of agreement, or range of proportions of agreement were more meaningful to report.

Statistical Methods

When comparing the selected N-MSD diagnostic label, perfect agreement and near-perfect agreement were defined. If the first label chosen by a PT matched the first or second label chosen by the GS, the agreement was coded as perfect agreement. If a second label was selected by a PT and it matched the first or second label of the GS, the agreement was considered a near-perfect.

Examination of agreement in functional tasks assessment involved some data preparation. All individual items listed under a task on the assessment form, were binary coded into 1, if checked by the PTs as present, and 0 if left blank. The only exception was for the few items that were listed under the essential component phase. When reviewing the data, the research team decided not to include these few items at the top of each task sub-phases. Usually, absence or lack of a movement element is more significant for determining the movement issues. The mere presence of essential components is believed to not necessarily be of value for diagnosing movement system deficits and only be used as a checklist when assessing a task. Thus, these items were left out of further analysis. Ultimately, the binary score of all individual items were added up to create summary task variables. Thus, this summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. Statistical analysis was done using the Statistical Package for the Social Sciences (SPSS) version 29. The assessment form data was transferred to the software.

4.3 RESULTS

The purpose of this study was to examine the agreement in ratings between two groups of PTs with different experience levels in analyzing different tasks based on the N-MSD framework and assigning N-MSD diagnostic labels in individuals with various neurological health conditions.

Fourteen patients with neurological conditions and 23 physical therapists were included in this study. Descriptive data for patients and PTs were summarized in Table 4.1. Previous knowledge and exposure of PTs to topics related to MSDs were shown in Table 4.2. Majority of the physical therapist participants (n=15) learned and followed the news about neurology-specific MSD frameworks within the past decade, including the publications by the ANPT MSD Task Forces, presentations and educational sessions at conferences, etc.

Table 4.1 Descriptives of Physical Therapist Participants

Measure	Total		Early Career PTs		Experienced PTs	
	n = 23	*%*	*n = 12*	*%*	*n = 11*	*%*
Demographics						
*Employment Status **						
Full time PT job	18	78.3	12	100%	6	54.5
Research, Teaching or Part time PT job	7	30.4	0	0	7	30.4
*Settings of practice ***						
Inpatient (Hospital)	8	35.6	4	33.3	4	36.4
Outpatient/Private Practice	6	26.1	2	8.7	4	36.4
Neuro Case workload						
70% and more	18	78.3	9	75	9	81.8
Neurologic Clinical Specialist (NCS)	11	47.8	4	33.3	7	63.6

Note. PT = Physical Therapist. **Participants were allowed to choose more than one option.*
***Participants who worked only at one of the two settings.*

Table 4.2 Descriptives of Patient Participants

Measure	*n=14*	*%*
Demographics		
Male	6	43
Female	8	57
Age (M, SD)	62.72	17.43
Time since Diagnosis (M, SD)	13.14	9.31
Use of Assistive Device	5	36
Health Condition		
PD	5	36
MS	3	21
CVA	2	14
Mix	4	29
N-MSD Dx		
1.Movement Pattern Coordination Deficit (MPCD)	4	29
2.Force Production Deficit (FPD)	5	36
3.Fractionated Movement Deficit (FMD)	3	20
4.Postural Vertical Deficit (PVD)	0	0
5.Sensory Selection & Weighting Deficit (SSWD)	1	7
6.Sensory Detection Deficit (SDD)	0	0
7.Dysmetria (DM)	1	7
8.Hypokinesia (HK)	0	0

Note. N-MSD = Neurologic Movement System Diagnosis, CVA = Cerebrovascular Accident, PD= Parkinson's Disease, MS= Multiple Sclerosis.

Table 4.3 Descriptives of MSD knowledge of Physical Therapist participants

Measure	Total		Early Career PTs		Experienced PTs	
	n = 23	*%*	*n = 12*	*%*	*n = 11*	*%*
MSD knowledge						
Not at all	2	8.7	2	16.7	0	0
I have heard about the concept, but I don't know about any particular system	3	13	2	16.7	1	9.1
I am familiar with the concept, and have heard about couple of systems or diagnostic labels, but never used any in practice	5	21.7	4	33.3	1	9.1
I am familiar with the concept and examples, and have used some neuro systems/labels in my practice, occasionally as they fit	7	30.4	2	16.7	5	45.5
I am familiar with the concept and examples and use neuro systems/labels often in practice	6	26.1	2	16.7	4	36.4
Introduction to MSD						
I learned about the concept decades ago when it was first introduced (~1990's)	1	4.3	0	0	1	9.1
I learned about specific orthopedic systems in 2000's	1	4.3	0	0	1	9.1
I learned about the concept within the past decade	8	34.8	4	33.3	4	36.4
I have followed the publications and correspondences on MSD since 2015	7	30.4	0	0	7	63.6
I have read the 2017 APTA White Paper: "Movement System Diagnoses in Neurologic Physical Therapy"	15	65.2	5	41.7	10	90.9

I have attended presentations and talks at conferences, educational settings, at work, etc.	12	52.2	4	33.3	8	72.7
I have taken courses/attended workshops on a specific orthopedic system	2	8.7	0	0	2	18.2
I have taken courses/attended workshops on a specific neurological system	6	26.1	2	16.7	4	36.4
N-MSD Knowledge						
Are you familiar with this system?	15	65.2	5	41.7	10	90.9
I have attended presentations and talks at conferences, educational settings, at work, etc.	8	34.8	3	25	5	45.5
I have taken courses/attended workshops on this system	2	8.7	0	0	2	18.2
This system was thought in my DPT/Residency program	3	13	1	16.7	2	9.1
I learned about this system when the Academy of Neurologic Physical Therapy (ANPT) developed the Movement System Task Force in 2015	12	52.2	2	16.7	10	90.9
I have read the papers published by the Movement System Task Force in 2021	12	52.2	3	25	9	81.8
Use of N-MSD in clinical practice						
I don't currently use it	3	13	2	16.7	1	9.1
I use this system occasionally in my clinical practice	2	8.7	1	8.3	1	9.1
I use this system routinely in my clinical practice	7	30.4	2	16.7	5	45.5
I am teaching this content in a DPT/residency program	8	34.8	0	0	8	72.7
I am involved with research in this area	2	8.7	0	0	2	18.2

Note. PT = Physical Therapist.

Agreement in N-MSD label assignment:

To examine the agreement rate between early career and experienced therapists, first their performance was compared against GS (table 3) in assignment of the N-MSD labels. The overall agreement rate between PTs and GS was moderate (40%). However, when PTs were grouped based on their experience, the agreement rate was noticeably different. Overall, early-career PTs agreed with the GS in 30 percent of the assessments (8 out of 27), whereas experienced PTs were in agreement with the GS in 52% of the times (13 out of 25 assessments). When comparing the two groups of PTs, those with more experience were in perfect agreement with the GS significantly more times than PTs in early-career stages ($X^2(1)= 6.71$, $p = 0.01$). Additionally, the agreement of assessments done by the early-career PTs were equally divided between perfect and near-perfect estimates (4/27, 15% each).

A subgroup of experienced PTs (n=5) were current or former members of various ANPT Taskforces that were primarily focused on development and application of movement system diagnosis in balance dysfunctions or movement analysis of task frameworks. The 13 assessments done by these 5 individuals showed an agreement rate of 65% (9/13) between PTs and the GS. When only accounting for individuals involved with diagnosis development and application, the agreement rate with the GS raised to 82% (9/11) on 7 patients with PD, MS, CVA and mix neurologic (ALS) diagnoses.

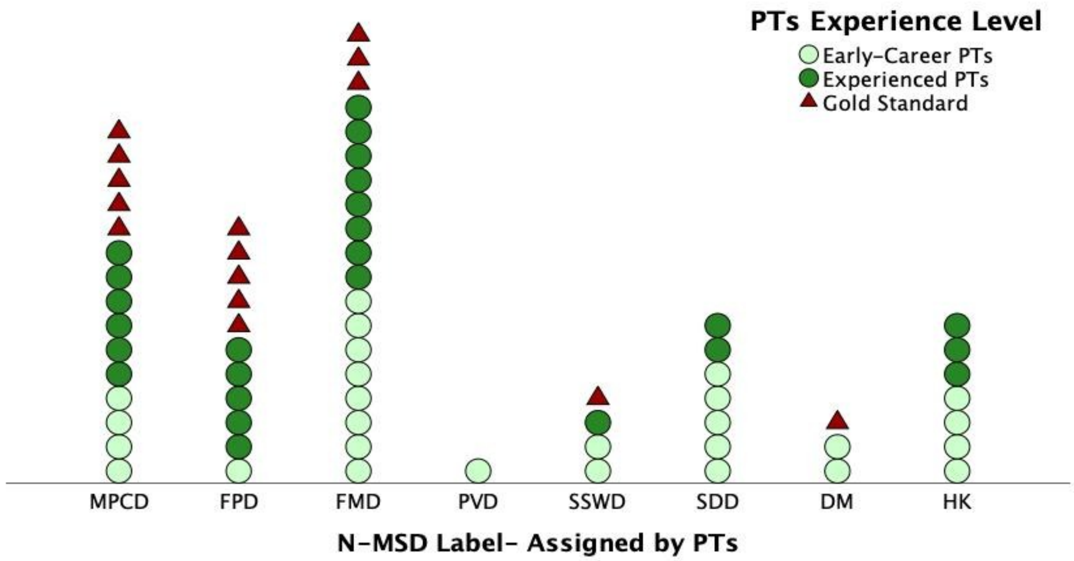
Table 4.4 Agreement Between Physical Therapists and Gold Standard in Assigning N-MSD Label per Experience level

	Perfect Agreement		Near-Perfect Agreement		Overall Agreement	
	Proportion	Percentage	Proportion	Percentage	Proportion	Percentage
All therapists (n=23)	16/52	0.31	5/52	0.09	21/52	0.40
Agreement per PTs experience level						
Early career (patients=13, PTs=12)	4/27	0.15	4/27	0.15	8/27	0.30
Experienced (patients=14, PTs=11)	12/25	0.44	1/25	0.08	13/25	0.52
Expert (patients=7, PTs=5)	8/13	0.62	1/13	0.08	9/13	0.69

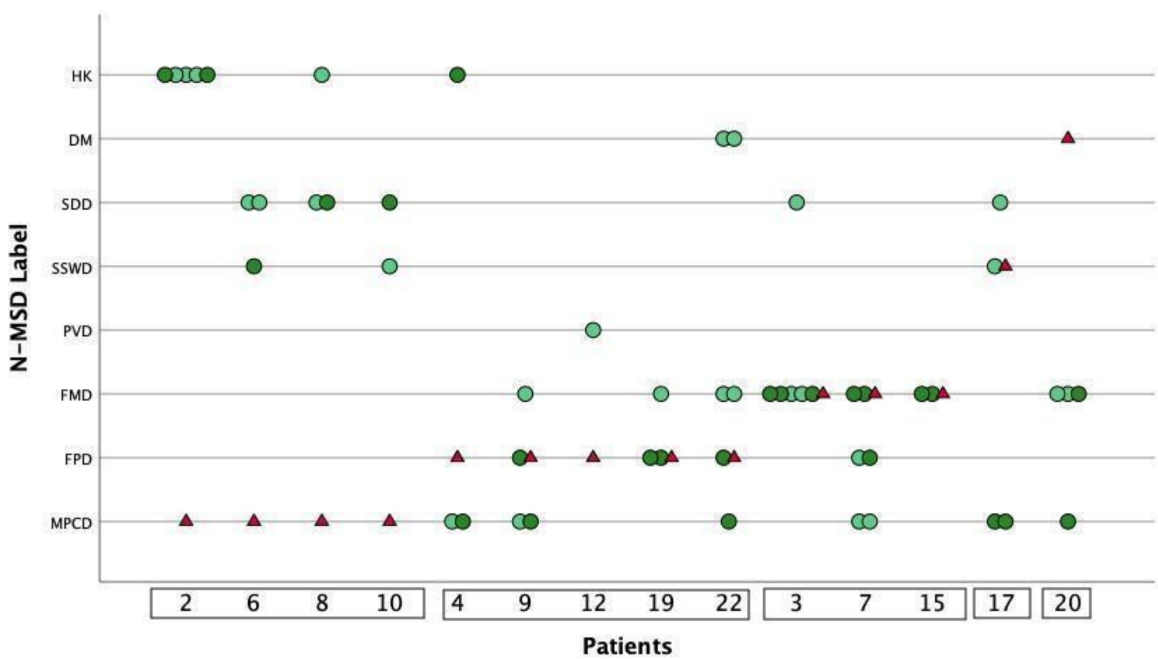
Note. Each PT assessed at least 2 patients. Each patient was seen by at least 2 PTs (except pt 12)
CVA: Cerebrovascular Accident, PD: Parkinson's Disease, MS: Multiple Sclerosis
MPCD: Movement Pattern Coordination Deficit, FPD: Force Production Deficit, FMD: Fractionated Movement Deficit, SSWD: Sensory Selection and Weighting Deficit, DM: Dysmetria.

The label assignment spread was depicted in Graph 4.1. Each dot represents an assessment that was done by a PT. The GS assessments on 14 patients are depicted in red. Bright green shows the assessments done by early-career PTs and dark green depicts ratings by experienced PTs. The Graph 4.1 (a) shows the number of times that each diagnosis was chosen across all patients. Graph 4.1 (b) shows the diagnoses that were chosen by PTs for each patient. The stacked dots show when more than one PT chose a label. The red triangle shows the chosen diagnosis by the GS.

When grouping assessments based on the N-MSD label assigned by the GS, PTs seemed to agree with each other as well as the GS more readily for the FMD category compared to other labels. Assessing patients identified as MPCD by the GS seemed to be the hardest for PTs in both groups as their chosen labels were spread out across three labels, none of which matched the GS label. In several cases there were at least a pair of PTs from the same group agreeing on the same label. Early-career PTs agreed with each other on patients 6, 22, and 7, while their experienced peers chose different labels. Experienced PTs agreed with each other on patients 19, 7 and 17. However, these between PT group differences did not follow a distinguishable pattern in terms of patient presentation or movement impairments.



Graph 4.1 (a) N-MSD Label Assignment by Physical Therapists per Movement Impairment

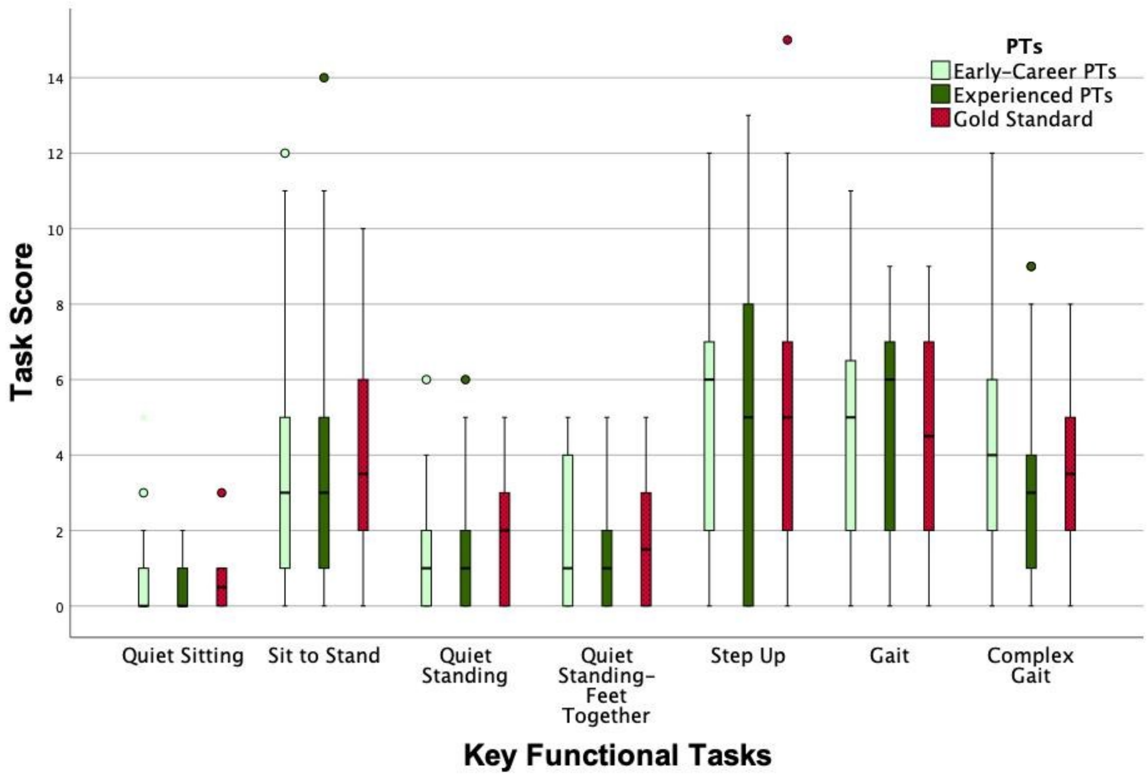


Graph 4.1 (b) N-MSD Label Assignment by Physical Therapists per Patients

1.MPCD: Movement Pattern Coordination Deficit, 2.FPD: Force Production Deficit, 3.FMD: Fractionated Movement Deficit, 4.PVD: Postural Vertical Deficit, 5.SSWD: Sensory Selection and Weighting Deficit, 6.SDD: Sensory Detection Deficit, 7.D: Dysmetria, 8.H: Hypokinesia.

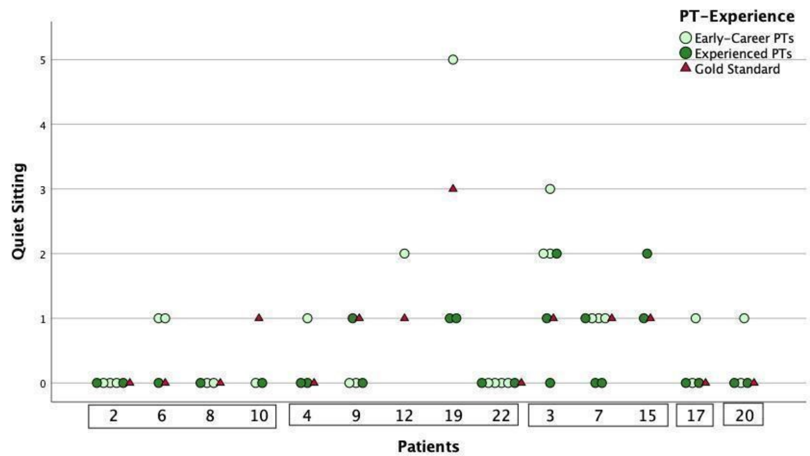
Agreement in N-MSD task assessment:

To examine the degree to which ratings were identical on task assessments between two groups of PTs, the task summary scores were used. As mentioned before, the task summary score was the sum of positive signs identified during each task, with higher scores indicating more observed movement deficits than lower scores. Graph 4.2 shows the task summary score analysis across all patients for both PT groups as well as the GS. The GS assessments of 14 patients showed that overall, the range and spread of task scores were greatest for Sit to Stand and Step Up tasks. Gait and Complex Gait tasks showed moderate range and variation in scores among patients while Quiet Sitting and Standing were scored with fewer impairments on the assessment forms. Assessments of the same 14 patients showed a similar pattern in that more impairments were identified during Sit to Stand and Step Up. Early-career PTs selected more impairments while assessing Quiet Standing-feet together and Complex Gait with median of their scores only one point or less higher than the experienced group.

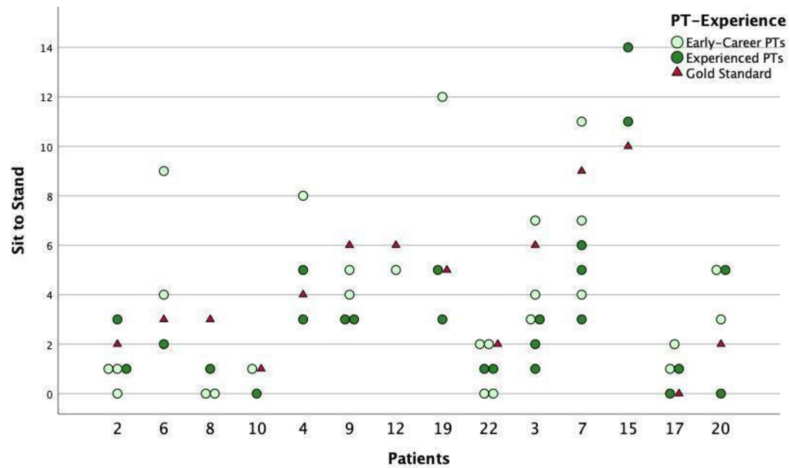


Graph 4.2 Task Assessments by Physical Therapists and the GS
PT scores on this graph accounts for both between patient variation in task scores, as well as between PTs scoring differences.

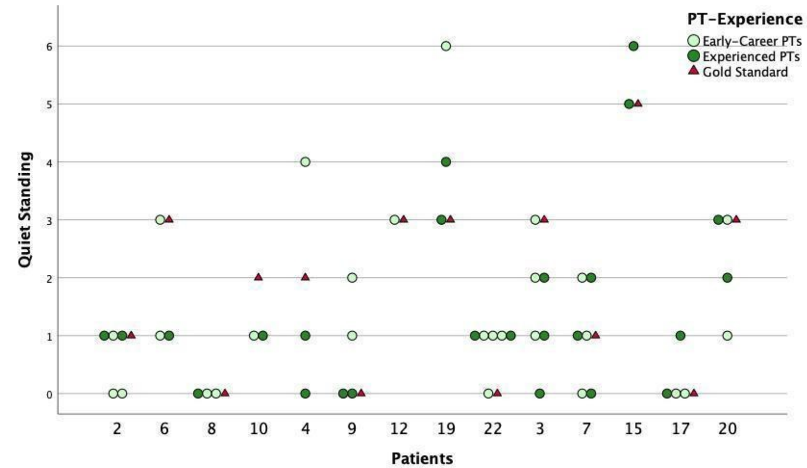
To differentiate the between-PTs variances in scores from between-patient variances, a case-by-case analysis was done (Graph 4.3). Each dot represents a task score that was given by a PT for a patient, across all seven core tasks. There were some general differences between how PTs scored different groups of patients. For example, patients with MPCD in comparison to other movement impairment groups, received lower scores (fewer deficits) on Sit to Stand and Step Up tasks. However, no consistent or obvious difference was found between ratings of early-career compared to experienced PTs.



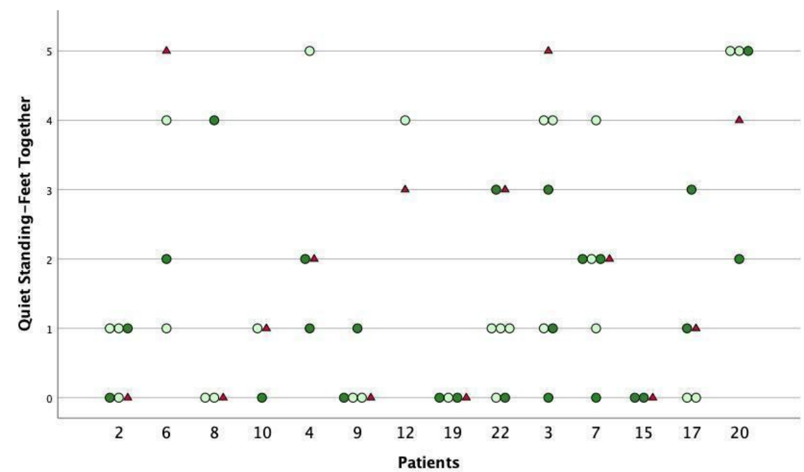
Graph 4.3 (a). Quiet Sitting



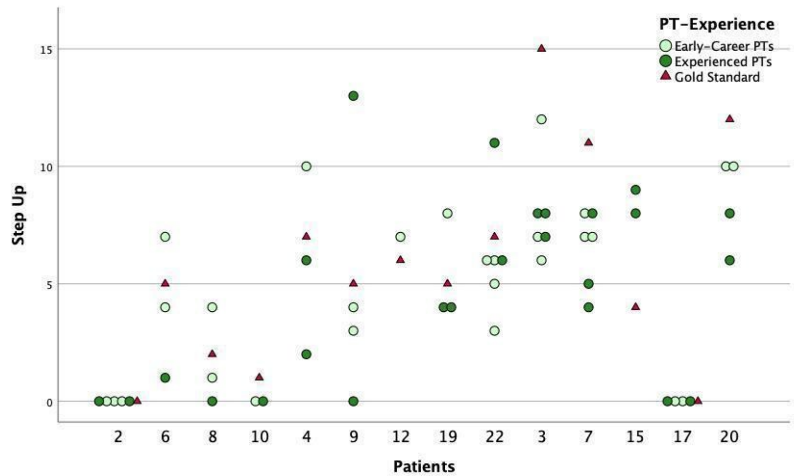
Graph 4.3 (b). Sit to Stand



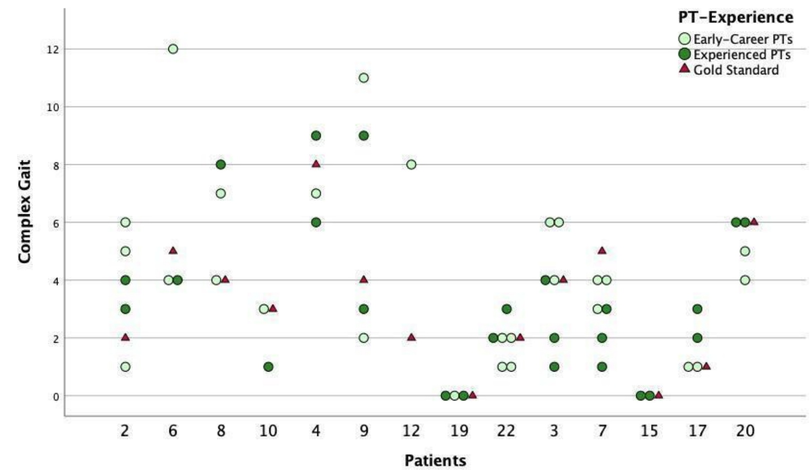
Graph 4.3 (c). Quiet Standing



Graph 4.3 (d). Quiet Standing-Feet Together

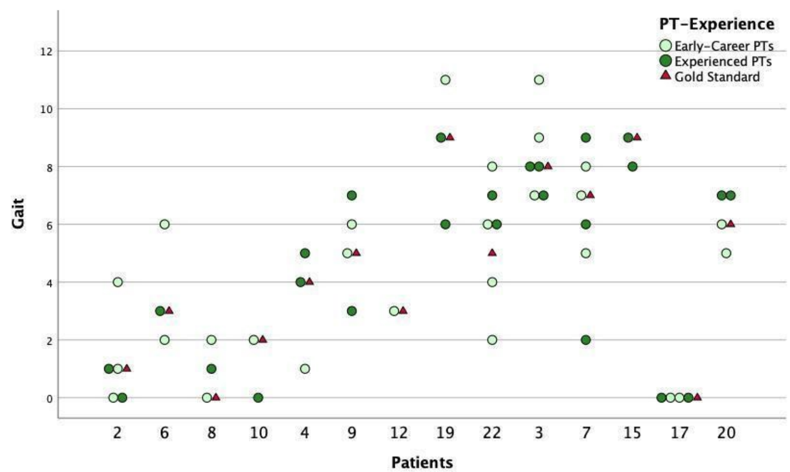


Graph 4.3 (e). Step Up



Graph 4.3 (g). Complex Gait

Graph 4.3 Task Assessments by Physical Therapists and the GS Per Patient



Graph 4.3 (f). Gait

4.4 DISCUSSION

A prerequisite for using a Movement System classification framework is for therapists to be able to reliably examine functional skills and group patients into the suggested movement deficit categories. The purpose of this study was to examine the agreement among PTs in using the N-MSD framework in patients with neurologic health conditions.

The label assignment analysis showed that therapists with different levels of clinical experience and previous familiarity with Movement System frameworks could assign the same diagnosis to patients as the GS, with moderate level of agreement (40%). The rate of agreement was higher for more experienced PTs, with some previous exposure or use of one MSD (52%) compared to the early career group (30%). Furthermore, the agreement between experienced PTs who were involved in ANPT Task Forces and the GS was as high as 65% in assigning N-MSD labels. These results suggested that more years of clinical experience and exposure and use of N-MSD labels could improve the agreement between PTs and experts when assessing patients. This is similar to the pattern found in reliability studies of other Movement System frameworks, such as the Movement System Impairment (MSI) classification system. In the MSI case, studies that rigorously examined the validity and reliability of the system were initially conducted by the developers (Van Dillen et al., 1998). When therapists outside of the MSI developing team were recruited for a reliability study, it was shown that with appropriate training, therapists could reliably classify people with low back pain (LBP). Later, independent researchers examined the reliability of the region-specific MSI categories by using the system in clinical settings and patient populations (knee pain, hip pain, lower back pain, etc.) (Harris-Hayes & Van Dillen, 2009). The reliability estimates improved gradually in studies that included PTs outside of the research team but properly trained.

A further analysis of the label assignment showed that therapists with different levels of clinical experience and previous familiarity with Movement System frameworks could assign the same diagnosis to patients with the same primary movement deficits, with an agreement rate of 69% to 82%. However, when the same patients were assessed by an expert, often a different N-MSD label was assigned to the group of patients. For example, patients that were identified with MPCD by an expert received Hypokinesia, SSD, and SSWD labels by the PT participants. This difference could be partially explained by how therapists usually associate a diagnosis with a certain medical condition. For example, patients with PD were often labeled with Hypokinesia as the main movement issue, while the same patient was identified by MPCD or FPD by the expert. Difficulty with the MPCD diagnosis could also be explained by the significant importance of the movement observations made during the Sit to Stand and Step Up task when identifying this condition. The greater variability of observations during these tasks may have contributed to more variability when selecting the MPCD diagnosis.

The training in this study included self-study of online, self-paced modules, with case examples for each diagnosis. While there were knowledge checkpoints integrated into the training modules to help with the learning process, there was no official monitoring of training completion by the PT participants. Therefore, the variance in agreement with the GS rates among even more experienced PTs could point to the need for a more rigorous training prior to the use of N-MSD.

Aside from the label assignment analysis, agreement in task assessments was also examined in this study. Analysis of the overall task scores showed that the stationary tasks such as Quiet Sitting and Standing consistently were marked with fewer deficits present compared to transitional and dynamic tasks such as Sit to Stand, Step Up and Gait. This finding was well explained by the fact that by design, there were fewer items listed for Quiet Sitting (6 deficits), Quiet Standing (9 deficits) to begin with. These static tasks are simpler tasks as compared to the transitional tasks. Moreover, the relative smaller variability in rating of these two tasks and therefore higher agreement compared to other tasks were partially due to the included participants. The average age of patient participants was 63 years, with 14 years since diagnosis on average. Besides the long-term condition, most of the patients had a high level of functionality, with only five using assistive devices. Therefore, lack of functional deficits during stationary tasks was an expected finding.

Analysis of the agreement among PTs in assessment of transitional and dynamic tasks did not show a consistent pattern, as agreement was high for some groups and low for others. In two occasions, four PTs assessed the only patient falling under a primary movement impairment (one patient with DM and one patient with SSWD labels). Given that there was no between-patient variance among the ratings in either of the groups, the within-patient rating variability was purely due to differences in PT's scoring of the patient. The PTs showed a relatively high level of agreement in scoring of all of the tasks (SD < 1.00), except for a couple of ratings. The agreement on the final N-MSD label was moderate in these cases (patient 17, 4 assessments: 50% agreement rate; patient 20, 4 assessments: 75% agreement rate). These patients were different in both level of function and severity of their conditions, as one was physically active in the community with minimum endurance impairments while the other one was limited in community walking heavily relying on another person close by.

The noticeable amount of variability in ratings of patients within a certain movement impairment group or a health condition often came down to a couple of "problematic" patients. For example, FPD and mix neurologic conditions groups showed the second most difference and dispersion of the ratings. However, patient by patient analysis showed that a single patient (patient 9) was the source for a significant amount of disagreement among PTs. One explanation for the discrepancy of ratings of this patient is that the neurological condition of ALS impacted the function of both sides. When assessing Step Up and Gait tasks, some PTs could have missed some deficits as they were considering one side as the more affected. It is possible that patients with a bilateral presentation in movement deficits, regardless of condition, may be more challenging to observe, may have more subtle findings, and therefore, may be more difficult to diagnose.

Limitations

This study had some limitations. The N-MSD framework was designed for PTs to do a thorough in-person physical examination of patients before making diagnostic decisions. In this study, PT participants were required to make assessments based on written impairment findings and short videos of patients performing functional tasks. While the GS for examining movement dysfunction is 3-dimensional (3-D) movement analysis, use of 2 dimensional (2-D) videos of patient performance that could be also replicated in clinical settings have been established as a reliable and validated method. Several studies provided evidence for moderate to high reliability for observational gait assessment through use of videotapes (Brunnekreef et al., 2005; Ferrarello et al., 2013). A study showed that by watching visual 2- D videos of athletes performing single leg squat, novice and experienced PTs were able to classify lower extremity movement patterns in patients with musculoskeletal disorders (Harris-Hayes et al., 2014).

The examination process and commands used in the patient functional task assessment were not standardized in our study. The examiner relied on the N-MSD manual for required adjustments, facilitation, and commands during tasks performances, however, there were inconsistencies in execution of the tasks. Partially caused by space, time, and patient ability constraints, the examiner’s discretion about modifications to the procedures may have introduced variations in task performances. Moreover, when reviewed by the GS, several assessments needed more in-depth investigation of findings that the examiner could have followed up with further testing.

The patient participants in this study were recruited from outpatient physical therapy clinics in the community. Hence, most of the participants were community-dwelling older adults capable of ambulating independently or with or without use of assistive devices or some assistance. This made the study sample limited in diversity in terms of individuals' function level, with no severe or acute case representation.

Gold standard reliability measures such as Kappa and Interclass Correlation Coefficients (ICCs) were not measured in this study due to the small sample size. Specific statistical approaches for assessment of reliability and agreement have been developed within different disciplines. Thus, there is no single statistical method to consider widely accepted to follow. It is also encouraged to report a combination of estimates when reporting the results to represent different statistical methods (Kottner et al., 2011). Nevertheless, the agreement proportions, percentages and range, although easily understandable, are limited to be extended for reliability purposes, as they do not adjust for agreements by chance. Other forms of reliability measures such as test-retest reliability are also recommended to be studied in future studies.

4.5 CONCLUSION

The use of movement-based diagnosis in physical therapy is becoming more prevalent in physical therapy. This study examined the agreement between early-career and experienced PTs in using the N-MSD framework. Our findings were in line with what was previously shown in the literature regarding the higher agreement among PTs with more experience. Those with expertise in neurologic physical therapy and familiarity with MS frameworks were in higher agreement with GS compared to early-career PTs. To integrate this framework into clinical practice, the required next step is to establish the psychometric properties of this movement system. With standardized techniques and proper training, early-career and experienced PTs could assess functional tasks and make movement-focused diagnoses.

Chapter 5. Conclusions

The work presented in this dissertation study focused on agreement among PTs in analyzing different tasks based on the N-MSD and assigning diagnostic labels in individuals with various neurological health conditions. The included studies aimed to examine the consensus among PTs with different clinical experiences and familiarity with the N-MSD framework in using detailed and comprehensive assessment forms when performing clinical observational analysis. Findings of this body of work demonstrated that PTs with more years of clinical practice, previous experience of using movement system diagnosis and appropriate training could reliably use N-MSD system and agree on the same diagnostic movement impairment label.

In chapter 2, a study was conducted to examine the level and extent of agreement between PTs and an expert in using the N-MSD. The study found a moderate overall agreement (40%) between PTs and the GS, with agreement improving significantly for experienced PTs (52%-65%) familiar with N-MSD. Training and prior exposure to movement system diagnosis were identified as crucial for consistency. Variability in task assessments, particularly transitional movements, indicated challenges in standardized evaluations. Considering the limitations in sample diversity and the need for rigorous statistical measures for reliability validation, future research studies are needed for more concrete task assessments among a more experienced and trained clinician.

In chapter 3, rate and proportion of agreement among PTs was examined in assigning N-MSD labels to patients with various neurologic health conditions. PTs showed moderate agreement (69%-82%), influenced by experience and familiarity with the framework. Stationary tasks like Quiet Sitting and Standing had higher consistency in ratings, while transitional and dynamic tasks, such as Step Up and Gait, revealed more variability, particularly for complex cases. Variations were linked to differing interpretations, patient complexity, and non-standardized task protocols. The findings highlight the need for standardized training and validated tools to enhance diagnostic reliability.

In chapter 4, rate and proportion of agreement between early-career and experienced PTs in assigning N-MSD labels to patients with various neurologic health conditions was examined. The agreement with the GS for all therapists was moderate (40%). Agreement improved with PTs' clinical experience and familiarity with N-MSD, rising to 52% for experienced therapists and 65% for those actively using or teaching the framework. Training quality emerged as a key factor, as rigorous training correlated with higher agreement rates. Task assessment analysis showed higher agreement in stationary tasks, likely due to fewer deficits and variability, while transitional and dynamic tasks showed inconsistent agreement. Variability often stemmed from individual patients, with notable discrepancies linked to complex cases like ALS or differing functional levels. The findings of this study highlighted the importance of training and clinical experience in enhancing PTs' reliability when using movement system frameworks.

The collective findings of this body of work highlight several important directions for future research. To strengthen and validate these results, subsequent studies should include larger and more diverse patient samples, particularly encompassing a wider range of functional abilities and movement impairments. Additionally, it is essential to include a sample of expert PTs, with extensive clinical experience, who actively contribute to the ongoing development of this field through participation in task forces and professional initiatives. The inclusion of such experts in future assessments would provide valuable insights into key decision making parameters used by this group that could be incorporated into the N-MSD guidelines/manual.

Moreover, the findings of this study emphasize the need for a more standardized and unified framework for observational task analysis. While clinicians appear to demonstrate stronger agreement on assessment of certain movement components, there is notably less consensus regarding observational assessment of transitional tasks, Sit to Stand and Step Up. This variability underscores the necessity of developing systematic tools for the assessment of transitional tasks, which would support both research efforts and the integration of these skills into educational curricula for training clinicians.

Finally, the results suggest that with comprehensive training in this Movement System framework, moderate to high levels of inter-rater agreement are achievable. Establishing greater consensus on diagnostic classifications is critical, as these shared understandings directly inform clinical decision-making, including treatment strategies, plans of care, and patient prognosis. Continued refinement of assessment methods and educational approaches will therefore be essential to advancing both the research and clinical application of this framework, ultimately contributing to improved patient outcomes.

Bibliography

Allen, D. D. (2007). Proposing 6 dimensions within the construct of movement in the movement continuum theory. *Physical Therapy*, *87*(7), 888–898. https://doi.org/10.2522/ptj.20060182

Brunnekreef, J. J., van Uden, C. J., van Moorsel, S., & Kooloos, J. G. (2005). Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC Musculoskeletal Disorders*, *6*(1), 17. https://doi.org/10.1186/1471-2474-6-17

Brunnstrom, S. (1970). *Movement therapy in hemiplegia: A neurophysiological approach.* ([1st ed.]). New York, Medical Dept., Harper & Row.

Caldwell, C., Sahrmann, S., & Van Dillen, L. (2007). Use of a movement system impairment diagnosis for physical therapy in the management of a patient with shoulder pain. *The Journal of Orthopaedic and Sports Physical Therapy*, *37*(9), 551–563. https://doi.org/10.2519/jospt.2007.2283

Chua, C. L., Chong, R., Hill, S., Ng, S., Chan, J., Hassan, N., Patterson, K., & Vaughan-Graham, J. (2024). *Observational movement analysis and clinical reasoning about the sit to stand movement in a person with stroke: Exploring perspectives of student and expert neurological physiotherapists*. https://www.authorea.com/doi/full/10.22541/au.170670363.38189986/v1?commit=23c4bf5d983003f3809dcc8161f8a19aa85cbed6

Coffin-Zadai, C. A. (2007). Disabling Our Diagnostic Dilemmas. *Physical Therapy*, *87*(6), 641–653. https://doi.org/10.2522/ptj.20060236

Cools, A. M., & Michener, L. A. (2017). Shoulder pain: Can one label satisfy everyone and everything? *British Journal of Sports Medicine*, *51*(5), 416–417. https://doi.org/10.1136/bjsports-2016-096772

Deutsch, J. E., Gill-Body, K. M., & Schenkman, M. (2022). Updated Integrated Framework for Making Clinical Decisions Across the Lifespan and Health Conditions. *Physical Therapy*, *102*(3), pzab281. https://doi.org/10.1093/ptj/pzab281

Dobkin, B. H. (2009). Motor rehabilitation after stroke, traumatic brain, and spinal cord injury: Common denominators within recent clinical trials. *Current Opinion in Neurology*, *22*(6), 563–569. https://doi.org/10.1097/WCO.0b013e3283314b11

Eastlack, M. E., Arvidson, J., Snyder-Mackler, L., Danoff, J. V., & McGarvey, C. L. (1991). Interrater reliability of videotaped observational gait-analysis assessments. *Physical Therapy*, *71*(6), 465–472. https://doi.org/10.1093/ptj/71.6.465

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of Surgical Oncology*, *17*(6), 1471–1474. https://doi.org/10.1245/s10434-010-0985-4

Ferrarello, F., Bianchi, V. A. M., Baccini, M., Rubbieri, G., Mossello, E., Cavallini, M. C., Marchionni, N., & Di Bari, M. (2013). Tools for Observational Gait Analysis in Patients With Stroke: A Systematic Review. *Physical Therapy*, *93*(12), 1673–1685. https://doi.org/10.2522/ptj.20120344

Gor-García-Fogeda, M. D., Cano de la Cuerda, R., Carratalá Tejada, M., Alguacil-Diego, I. M., & Molina-Rueda, F. (2016). Observational Gait Assessments in People With

Neurological Disorders: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*, *97*(1), 131–140. https://doi.org/10.1016/j.apmr.2015.07.018

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Harris-Hayes, M., Steger-May, K., Koh, C., Royer, N. K., Graci, V., & Salsich, G. B. (2014). Classification of Lower Extremity Movement Patterns Based on Visual Assessment: Reliability and Correlation With 2-Dimensional Video Analysis. *Journal of Athletic Training*, *49*(3), 304–310. https://doi.org/10.4085/1062-6050-49.2.21

Harris-Hayes, M., & Van Dillen, L. R. (2009). The Inter-Tester Reliability of Physical Therapists Classifying Low Back Pain Problems Based on the Movement System Impairment Classification System. *PM&R*, *1*(2), 117–126. https://doi.org/10.1016/j.pmrj.2008.08.001

Hedman, L. D., Quinn, L., Gill-Body, K., Brown, D. A., Quiben, M., Riley, N., & Scheets, P. L. (2018). White Paper: Movement System Diagnoses in Neurologic Physical Therapy. *Journal of Neurologic Physical Therapy*, *42*(2), 110–117. https://doi.org/10.1097/NPT.0000000000000215

Hedman, L. D., Rogers, M. W., & Hanke, T. A. (1996). Neurologic Professional Education: Linking the Foundation Science of Motor Control With Physical Therapy Interventions for Movement Dysfunction. *Journal of Neurologic Physical Therapy*, *20*(1), 9.

Hornby, T. G., Reisman, D. S., Ward, I. G., Scheets, P. L., Miller, A., Haddad, D., Fox, E. J., Fritz, N. E., Hawkins, K., Henderson, C. E., Hendron, K. L., Holleran, C. L., Lynskey, J. E., & Walter, A. (2020). Clinical Practice Guideline to Improve Locomotor Function Following Chronic Stroke, Incomplete Spinal Cord Injury, and Brain Injury. *Journal of Neurologic Physical Therapy*, *44*(1), 49. https://doi.org/10.1097/NPT.0000000000000303

Hulleck, A. A., Menoth Mohan, D., Abdallah, N., El Rich, M., & Khalaf, K. (2022). Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. *Frontiers in Medical Technology*, *4*, 901331. https://doi.org/10.3389/fmedt.2022.901331

Jensen, G. M., Gwyer, J., Shepard, K. F., & Hack, L. M. (2000). Expert Practice in Physical Therapy. *Physical Therapy*, *80*(1), 28–43. https://doi.org/10.1093/ptj/80.1.28

Jensen, G. M., Shepard, K. F., & Hack, L. M. (1990). The Novice Versus the Experienced Clinician: Insights into the Work of the Physical Therapist. *Physical Therapy*, *70*(5), 314–323. https://doi.org/10.1093/ptj/70.5.314

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, *64*(1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002

Ludewig, P. M., Kamonseki, D. H., Staker, J. L., Lawrence, R. L., Camargo, P. R., & Braman, J. P. (2017). CHANGING OUR DIAGNOSTIC PARADIGM: MOVEMENT SYSTEM

DIAGNOSTIC CLASSIFICATION. *International Journal of Sports Physical Therapy*, *12*(6), 884–893.
McClure, P. W., & Michener, L. A. (2015). Staged Approach for Rehabilitation Classification: Shoulder Disorders (STAR–Shoulder). *Physical Therapy*, *95*(5), 791–800. https://doi.org/10.2522/ptj.20140156
Observational Gait Analysis Handbook. (1989). Professional Staff Association, Rancho Los Amigos Medical Center.
Petraglia, F., Scarcella, L., Pedrazzi, G., Brancato, L., Puers, R., & Costantino, C. (2019). Inertial sensors versus standard systems in gait analysis: A systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, *55*(2), 265–280. https://doi.org/10.23736/S1973-9087.18.05306-6
Quinn, L., Riley, N., Tyrell, C. M., Judd, D. L., Gill-Body, K. M., Hedman, L. D., Packel, A., Brown, D. A., Nabar, N., & Scheets, P. (2021). A Framework for Movement Analysis of Tasks: Recommendations From the Academy of Neurologic Physical Therapy's Movement System Task Force. *Physical Therapy*, *101*(9), pzab154. https://doi.org/10.1093/ptj/pzab154
Ridao-Fernández, C., Pinero-Pinto, E., & Chamorro-Moriana, G. (2019). Observational Gait Assessment Scales in Patients with Walking Disorders: Systematic Review. *BioMed Research International*, *2019*(1), 2085039. https://doi.org/10.1155/2019/2085039
Rosenbaum, P., Eliasson, A.-C., Hidecker, M. J. C., & Palisano, R. J. (2014). Classification in Childhood Disability: Focusing on Function in the 21st Century. *Journal of Child Neurology*, *29*(8), 1036–1045. https://doi.org/10.1177/0883073814533008
Sahrmann, S. (2002). *Diagnosis and Treatment of Movement Impairment Syndromes*. Elsevier Health Sciences.
Sahrmann, S. A. (2014). The Human Movement System: Our Professional Identity. *Physical Therapy*, *94*(7), 1034–1042. https://doi.org/10.2522/ptj.20130319
Sahrmann, S., Azevedo, D. C., & Dillen, L. V. (2017). Diagnosis and treatment of movement system impairment syndromes. *Brazilian Journal of Physical Therapy*, *21*(6), 391–399. https://doi.org/10.1016/j.bjpt.2017.08.001
Saladin, L., & Voight, M. (2017). INTRODUCTION TO THE MOVEMENT SYSTEM AS THE FOUNDATION FOR PHYSICAL THERAPIST PRACTICE EDUCATION AND RESEARCH. *International Journal of Sports Physical Therapy*, *12*(6), 858–861.
Scheets, L. P., Hornby, T. G., Perry, S. B., Sparto, P., Riley, N., Romney, W., Fell, D., Kugler, K., & Nordahl, T. (2021). Moving Forward. *Journal of Neurologic Physical Therapy: JNPT*, *45*(1), 46–49. https://doi.org/10.1097/NPT.0000000000000337
Scheets, P. K., Sahrmann, S. A., & Norton, B. J. (1999). Diagnosis for Physical Therapy for Patients with Neuromuscular Conditions. *Journal of Neurologic Physical Therapy*, *23*(4), 158.
Scheets, P. L., Sahrmann, S. A., Norton, B. J., Stith, J. S., & Crowner, B. E. (2015). What is backward disequilibrium and how do i treat it? A complex patient case study. *Journal of Neurologic Physical Therapy: JNPT*, *39*(2), 119–126. https://doi.org/10.1097/NPT.0000000000000084
Scheets, Sahrmann, S. A., & Norton, B. J. (2007). Use of movement system diagnoses in the management of patients with neuromuscular conditions: A multiple-patient case report. *Physical Therapy*, *87*(6), 654–669. https://doi.org/10.2522/ptj.20050349

Van Dillen, L. R., Sahrmann, S. A., Norton, B. J., Caldwell, C. A., Fleming, D. A., McDonnell, M. K., & Woolsey, N. B. (1998). Reliability of physical examination items used for classification of patients with low back pain. *Physical Therapy*, *78*(9), 979–988.

Van Sant, A. F. (2017a). Movement System Diagnosis: *Journal of Neurologic Physical Therapy*, *41*, S10–S16. https://doi.org/10.1097/NPT.0000000000000152

Van Sant, A. F. (2017b). Movement System Diagnosis: *Journal of Neurologic Physical Therapy*, *41*, S10–S16. https://doi.org/10.1097/NPT.0000000000000152

Vision Statement for the Physical Therapy Profession. (2019, September 25). APTA. https://www.apta.org/apta-and-you/leadership-and-governance/policies/vision-statement-for-the-physical-therapy-profession

Wainwright, S. F., Shepard, K. F., Harman, L. B., & Stephens, J. (2010). Novice and Experienced Physical Therapist Clinicians: A Comparison of How Reflection Is Used to Inform the Clinical Decision-Making Process. *Physical Therapy*, *90*(1), 75–88. https://doi.org/10.2522/ptj.20090077

Wainwright, S. F., Shepard, K. F., Harman, L. B., & Stephens, J. (2011). Factors That Influence the Clinical Decision Making of Novice and Experienced Physical Therapists. *Physical Therapy*, *91*(1), 87–101. https://doi.org/10.2522/ptj.20100161

Witten, A., Mikkelsen, K., Mayntzhusen, T. W., Clausen, M. B., Thorborg, K., Hölmich, P., & Barfod, K. W. (2023). Terminology and diagnostic criteria used in studies investigating patients with subacromial pain syndrome from 1972 to 2019: A scoping review. *British Journal of Sports Medicine*, *57*(13), 864–871. https://doi.org/10.1136/bjsports-2022-106340

World Health Organization, O. (2001). *International classification of functioning, disability and health: ICF*. World Health Organization. https://iris.who.int/handle/10665/42407