ANGLIA RUSKIN UNIVERSITY

THE POTENTIAL YIELD OF EXTERNAL FEEDBACK VIA MOVEMENT SONIFICATION IN PHYSIOTHERAPY

JOÃO GUERRA

A thesis in partial fulfilment of the requirements of Anglia Ruskin University for the degree of Doctor of Philosophy

ACKNOWLEDGMENTS

DEDICATION

To my parents,

for their unconditional love and support

Thank you for giving me so much,

even when there wasn´t a lot.

but particularly,

to you Joana for being there,

while I crossed the endless darkness,

through the challenging storms,

couldn't have made it without you.

To my kids,

Make the world a better place,

dream big, believe yourselves and you can set the world alight.

"Only those who risk going too far can possibly know how far one can go"

T.S. Eliot

ABSTRACT

Objectives: New strategies to improve the validity, reliability and responsiveness of strength assessment in physiotherapy are needed. The purpose of this thesis was to develop a new hand-held dynamometer using sonification as a type of external feedback to improve the current assessment of muscle strength via concentric manual muscle tests.

Method: A hand-held dynamometer was tested in healthy subjects in a within-subject design to assess several joints to establish validity, reliability, responsiveness of peak torque, angle of peak torque and angular impulse. The device was tested with groups of testers with different characteristics. Sonification of manual muscle tests was used by physiotherapists with different experience levels to assess its effects on reliability, responsiveness and force production. A focus group with experienced physiotherapists who had used the device was performed to explore their perspective in using the prototype and how to facilitate its implementation in a clinical setting.

Results: The prototype can retrieve reliable information regarding peak torque and angular impulse in both upper and lower limb joints by both experienced and inexperienced testers. The prototype demonstrated validity in retrieving peak torque data in upper and lower limbs joints but not angle of peak torque. Sonification generally improved reliability for most testers in both joints tested. An increase in force production in 50% of the inexperienced testers was also found. Physiotherapists were satisfied with the use of the prototype and suggested software and hardware improvements.

Conclusion: A feasible hand-held dynamometer for physiotherapy was developed using low-cost parts. Acceptable levels of validity and reliability in different joints and with testers with diverse experience levels were demonstrated. The use of sonification in manual muscle tests appears to benefit manual muscle tests positively and might be helpful in the future for research and training purposes.

Keywords: Manual muscle tests; Hand-held dynamometer; Sonification; Physiotherapy; Muscle strength

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ix

COPYRIGHT DECLARATION

RESEARCH OUTPUT

Publications in peer-reviewed scientific journals

Guerra, J., Smith, L., Vicinanza, D., Stubbs, B., Veronese, N. and Williams, G., 2020. The use of sonification for physiotherapy in human movement tasks: A scoping review. *Science and Sports*. [online]

Research Award

Private Physiotherapy Educational Foundation. (2019). The potential yield of external feedback via movement sonification. João Guerra (Principal Investigator), L. Smith, D. Vicinanza and G. Williams. £14,768.03

LIST OF ABREVIATIONS

ACSM: American College of Sports Medicine

ASSA: Arduino-based Sound Strength Assessment

CI: Confidence Interval

ICC: Intra Class Correlation

IKD: Isokinetic Dynamometer

g: Grams

HHD: Hand-held Dynamometer

HSHPRC: Higher School of Health of The Portuguese Red Cross

Hz: Hertz

IMU: Inertial Measurement Unit

LOA: Limits of Agreement

MDC: Minimal Detectable Change

MDC%:   Minimal Detectable Change in percentage

MDD: Minimal Detectable Difference

MIDI: Musical Instrument Digital Interface

MMT: Manual Muscle Test

MVC: Maximal Voluntary Contraction

N: Newton

NA: Not applicable

Nm: Newton-meter

PCF: Participant Consent Form

PIS: Participant Information Sheet

PVC: Polyvinyl chloride

ROM: Range of Motion

SD: Standard Deviation

SEM: Standard Error of Measurement

v: Version

V: Volt

WLP: Weighted Least Products

1. INTRODUCTION

1.1 Research context

In general terms, data sonification is the use of audio signals to convey information or to increase data perceptibility (Hermann, 2008). It can be defined as a technique to transform data into sound so that the properties and relationships between data variables are communicated using audible signals and understandable by the listener while being systematic, objective, and reproducible (Hermann, 2008). This type of audio-feedback can communicate sounds' multiple variables, which can be manipulated to express quantitative parameters such as data from inertial measurement units (IMU) (Hermann, Hunt and Neuhoff, 2011). Sonification is a particularly powerful tool, as the auditory perception of extensive, complex, multidimensional data can be more refined than visual and statistical analyses. Sonification has been successfully employed in several scientific domains. Examples include the Geiger counter, in which the rate of clicking is a quantitative representation of the level of radiation near the device; the detection of plasma waves in planet atmospheres, whose oscillations are in the audio range (Scarf, et al., 1982); or its use as an alternative analytical method in DNA sequencing (Temple, 2017).

Sonification can also be conceptualised as a type of audio feedback that can support movement and assist exercise with or without supervision. This is done by using sound, which can enhance movement awareness and reproduction in rehabilitation tasks even if motor control or proprioception has not been affected while using a technological medium for data management. Technological solutions have enabled substantial progress and innovation in physiotherapy by supporting assessment and intervention (Russel, et al., 2011; Chanpimol, et al., 2017). For example, the use of motion analysis tools and audio-feedback systems to support physiotherapy has become more widespread (Knippenberg, et al., 2017; Sihvonen, et al., 2017; Whelan, et al., 2017).

The World Confederation of Physical Therapists defined physiotherapy as "services provided by physical therapists to individuals and populations to develop, maintain and restore maximum movement and functional ability throughout the lifespan" (WCPT, 2019). One of the main features of physiotherapy practice is its connection with movement and the ability to assess it, with the Health and Care Professions Council

15

stating in their standards of proficiency for physiotherapists the need to " be able to select and apply safe and effective physiotherapy specific practice skills including manual therapy, exercise and movement, electrotherapeutic modalities and kindred approaches" (HCPC, 2013). Muscular strength assessment falls within the knowledge physiotherapists need to possess in order to improve movement and function; however, its use can still be enhanced.

Technology can be explored in physiotherapy to improve the quality of service provided by growing information about patients or physiotherapy interventions. However, information is sometimes limited and qualitative. In physiotherapy, patients generally rely on proprioception, a mirror image, or on the therapist's advice; in those situations, other displays are difficult to integrate as the person moves freely in space (Großhauser, et al., 2012). Furthermore, during a physiotherapy session, the clinician often has limited access to quantitative information, such as applied forces, pressure, and weight distribution. This can be further explored in physiotherapy by using technology and sonification, which the author will demonstrate in this thesis.

1.2 Purpose

Strength testing is a general procedure within the physical examination. Strength assessment encompasses various techniques developed to evaluate force production on individual muscles or groups of muscles. There is an array of possibilities: from the gold-standard isokinetic dynamometer (IKD), to hand-held dynamometers (HHD), manual muscle tests and one-repetition maximum, among others. These approaches have been widely researched, and their limitations are widely known, with manual muscle tests the least discriminative in terms of information granularity. Nonetheless, manual muscle tests (in their various forms) are still commonly used as they have no cost and are easy to use. This thesis proposes a modification to the current approach to manual muscle tests by assessing concentric movements with a new and cheap HHD prototype.

Most societies in the world have been invaded by technology over the last few years. The widespread use of personal computers, mobile phones, the internet and the mass fabrication of its components has substantially reduced the cost of high-grade technological parts. Healthcare providers seem to be late adopters of technology, this was pointed by Poon, et al. (2006) in the United States, but also more recently in the United

Kingdom by Gardner, Webster and Barry (2018) when investigating the integration of regenerative medicine. This limitation impacts the quality of health outcomes, the cost and time-consumption of highly specialised human resources with Clark, et al. (2020) even proposing five points to facilitate technology integration in healthcare systems. As a part of most national health services (both public and private), this is also likely to be true in physiotherapy.

In this thesis, the author describes the development and testing of a new HHD prototype (ASSA – Arduino-based Sound Strength Assessment) that focuses on capturing peak force and angle of peak force across the range of movement in concentric manual muscle tests. This "dynamic" ability is a novel feature in HHDs that most common dynamometers do not possess. Moreover, this thesis assessed the viability of using concentric manual muscle tests for strength testing with and without sonification. In the context of this work, the sound feedback developed for the prototype - originated in the field of human-computer interaction - can be used to inform physiotherapist and/or patient on several parameters when a muscle test is underway, namely the speed at which the segment moves, the force applied throughout the range of motion, among others.

This work complements previous similar devices such as Li, et al. (2006) and Janssen and Le-Ngoc (2009) that have been investigated in manual muscle tests. At the same time, the thesis explores, for the first time, how sonification can be used in physical assessment. Sonification has been recently applied in motor learning and motor control tasks in several settings, by combining these two areas, it is possible to explore an area of knowledge where there is still room for expansion as technology makes its way into physiotherapy.

Below follows the thesis outline:

Chapter 1 – Introduction (page 15-20)

The current chapter provides an introduction to sonification and strength assessment in physiotherapy while framing the purpose of this research in light of the current framework.

Chapter 2 – Literature Review (page 21-41)

Chapter 2 provides a scoping review of the literature in which it was investigated how sonification has been used in the past for human movement tasks and how it can be applied for the benefit of physiotherapists. Gaps in the literature are identified to provide guidance for the development of the prototype.

A published version of this chapter can be found under the following citation:

Guerra, J., Smith, L., Vicinanza, D., Stubbs, B., Veronese, N. and Williams, G., 2020. The use of sonification for physiotherapy in human movement tasks: A scoping review. *Science and Sports*. [online]

Chapter 3 – Framework and device development (page 42-79)

Chapter 3 is divided into two sections; section 1 offers an overview of strength testing commonly used in physiotherapy. Section 2 focuses on device development and its various versions, followed by a sonification framework and an overview of the clinimetrics to be investigated.

Chapter 4 – Reliability and validity pilot test (page 80-123)

The pilot study described here assessed the prototype v2 reliability, responsiveness and validity when used by one physiotherapist as a tester and two non-physiotherapists in knee flexion and extension movements. The pilot test was also used to pre-test the study design and any flaws with the prototype.

Research question 1 – Is it feasible to assess strength using concentric manual muscle tests with a low-cost newly built prototype (ASSA v2)?

Aim 1 – Test the new prototype for reliability and validity in healthy subjects.

Chapter 5 – Reliability and validity with experienced physiotherapists (page 124 - 159)

This chapter describes the validity, reliability and responsiveness testing for the new prototype ASSA v3 with a group of experienced testers when assessing upper and lower limb movements with the new prototype in a group of healthy participants.

Research question 2 – Is ASSA v3 validity, reliability and responsiveness acceptable in assessing strength using concentric manual muscle tests with a low-cost newly built prototype when used by experienced testers?

Aim 2 – Test ASSA v3 for validity, reliability and responsiveness in healthy subjects with a homogenous group of experienced testers.

Chapter 6 – Sonification testing of reliability with experienced physiotherapists vs physiotherapy students (page 160-206)

A further update on the software demonstrates how the new ASSA v4 can be used for sonification and how it affects reliability in experienced and inexperienced users.

Research question 3 – Is ASSA v4 reliable and responsive in assessing strength using concentric manual muscle tests with a low-cost newly built prototype when used by both experienced and inexperienced testers?

Aim 3 – Test ASSA v3 for reliability and validity in healthy subjects with a heterogenous group of testers.

Research question 4 – Can sonification provide relevant feedback in concentric manual muscle tests when used by both experienced and inexperienced testers?

Aim 4 – Test ASSA v4 the ability of sonification to affect force production and reliability in healthy subjects with a heterogenous group of testers.

Chapter 7 – Focus group (page 207-223)

In chapter 7, physiotherapists who used the device provide an insight into the opinion from physiotherapists about using the new prototype, how it can be improved and how sonification can be implemented in clinical practice.

Chapter 8 – Discussion and conclusion (page 224-242)

The discussion provides a general overview of findings, where the authors frames the findings in relation to similar research while explaining the contribution to knowledge from both the development of the prototype and the major findings. A conclusion summarising the above is also provided.

2. LITERATURE REVIEW

There has been some work done overlapping the areas of physiotherapy and sonification, but before selecting the best approach for the current research, it was necessary to understand current developments in the area better. Therefore, to gather information and provide guidance, a scoping review was performed. Background information about sonification and its use in physiotherapy is thus provided here.

2.1 Introduction

The use of sonification in physiotherapy and rehabilitation is a relatively uncultivated domain. Although physiotherapists have used audio-feedback techniques in their intervention, it mainly focused on verbal or sound feedback during a particular task (ex.: verbally cueing a Parkinson's patient during gait training) and not necessarily the by-product of data from human movement transformed into a relevant and informative sound. In addition, auditory perception of complex, structured information could have several advantages in terms of temporal, amplitude, and frequency resolution when compared to visual representations and often opens up possibilities for an alternative or complement to visualisation techniques when available (Hermann, Hunt and Neuhoff, 2011). These advantages include the human ear's capability to detect patterns, recognise timbres, and follow different strands at the same time. In a natural way, this would offer the opportunity to render different, interdependent variables related to human movements in the sound in such a way that a listener (the patient or the therapist) could gain relevant insight into the represented information or data.

Sonification is not a novel feature in human-computer interaction but, with the current technological revolution, it has been using the immense depth of our auditive capabilities to develop an enlarged perspective on perceiving and analysing information. This is of particular importance due to the increased effortlessness to capture millions of variables on several fields, resulting in the creation of the so-called big data phenomenon. Sonification has also shown important effects on human movement capabilities, having been used in motor control research in the past. Therefore, exploring it within the field of physiotherapy and human movement was a logical next step.

The use of technology to capture quantitative data is now increasingly used in physiotherapy. For example, optical motion capture systems and portable IMU's can harness data from a moving segment or during a functional task. Characteristics of these movements can be communicated in the form of audio feedback delivered to both clinician and patient. This can provide insights into human movement that current approaches cannot capture, for example, by informing (through audio) on a joint range of movement, segment orientation, or muscle activation during a specific assessment or rehabilitation task. Therefore, sonification and its ability to take quantitative data and transform it into sound is becoming increasingly more relevant in physiotherapy.

The present scoping review aims to: (1) identify current research that has used sonification for physiotherapy or that has the potential to be used for physiotherapy (thus on populations who might undergo rehabilitation) in healthy individuals and/or those with a specific movement dysfunction; (2) identify technological approaches used to date, and functional tasks/movements investigated; (3) study sonification's potential effects on human subjects, particularly in motor control; (4) map literature pertaining to the current use of human movement sonification to identify gaps in the literature and directions for future research.

2.2 Method

Scoping reviews are a valuable approach to address the aim of this work, where charting current knowledge in an emergent area is the primary goal. Mainly, when trying to understand what the critical characteristics within a research field are, a scoping review allows researchers to identify gaps in the literature where methodological quality assessment is not mandatory - a characteristic of systematic reviews (Munn, et al., 2018).

The framework present in this scoping review was proposed by Arksey and O'Malley (2005). Several steps were followed to ascertain methodological consistency: "1. Identifying the research question, 2. Identifying relevant studies, 3. Study Selection, 4. Charting the data, 5. Collating, summarising and reporting the results", as stated by Arksey and O'Malley (2005).

The research question for this paper is: How has sonification been used in Physiotherapy for rehabilitation and to assist motor control?

The papers were categorised according to the study type, the technology used to capture human movement, the movement or tasks being performed, and the reported effects of sonification. It was decided to capture most advances in the field, to chart the data from all papers retrieved, regardless of methodological quality. Therefore, relevant information regarding new approaches in the area could be mapped for future reference.

Data from included studies were compiled on an Excel spreadsheet with information collected on the following topics: study aims, participant characterisation, activity investigated, the technique used for movement analysis, the sonification approach, and the effect on human movement.

2.2.1 Databases and Search Process

Databases searched: Web of Science, Science Direct, IEEE Explore, SportDiscus, Scopus and Pubmed. Each database was searched for the following keywords: "Sonification" AND "Physical therapy", "Sonification" AND "Physiotherapy", "Sonification" AND "Rehabilitation", "Sonification" AND "Motor Control", "Sonification" AND "Exercise Therapy". These keywords were selected to guarantee broad and relevant search results from Health and Technology related databases.

Keywords "rehabilitation" and "exercise therapy" were included to cover a broader area of movement science research to increase results regarding human movement sonification.

2.2.2 Eligibility

To maximise the number of results, any study that used sonification about human movement intending to develop or which could be used in physiotherapy practice was included. However, the technology had to be tested in human subjects and not only a prototyped idea under development. Object sonification and guidance tasks were excluded as these are usually more relevant for visually impaired people, although these can have relevance for some professionals.

The following inclusion and exclusion criteria were used:

- **Inclusion**: Sonification as an audio-feedback technique to provide information about human movement; sonification that has been tested in human subjects; peer-reviewed journals; books; thesis or conference presentations.

- **Exclusion**: Sonification of human movement that does not relate to physiotherapy, rehabilitation or motor control; object sonification; guidance tasks.

These criteria were selected to maximise the information collected, and there was no restriction on language, type of sonification, or subject.

2.2.3   Screening

The screening procedure follows the system suggested by Pham, et al. (2014). In order to select papers for review, the title was analysed for relevance and duplicates were excluded. Two researchers reviewed these papers to achieve uniformity in the inclusion criteria. The two authors reviewed 30% of the excluded papers and achieved 100% agreement on paper inclusion/exclusion criteria.

RefWorks Software was used to manage and exclude duplicates. The remaining papers had their abstract examined to meet the inclusion criteria. As per previous scoping reviews, selected papers were not reviewed for their methodological quality.

2.3 Results

The search covered literature published until the 3rd of January 2018. The total number of papers was: Web of Science - 57, Science Direct - 149, IEEE Explore - 111, SportDiscus - 6, Scopus - 702 and Pubmed - 33. The total number of papers identified from the search was 1058.

After duplicates and irrelevant titles were removed, 264 papers were left. All the abstracts of these papers were reviewed by the first reviewer, and a list of 73 papers was obtained based on the inclusion criteria, see figure 2.1.

Figure 2. 1 Flow chart of studies

```
┌─────────────────────────┐
│ Total records from database │
│ search (n=1058)         │
└─────────────────────────┘
            │
            ▼
┌─────────────────────┐        ┌──────────────────────────┐
│ Records screened    │──────▶ │ Duplicates and irrelevant titles │
│ (n=1058)            │        │ removed (n=794)          │
└─────────────────────┘        └──────────────────────────┘
            │
            ▼
┌─────────────────────┐        ┌──────────────────────────┐
│ Full-text articles  │        │ Records added from other sources │
│ assessed (n=73)     │        │ (n=13)                   │
└─────────────────────┘        └──────────────────────────┘
            │                              │
            ▼                              ▼
       ┌──────────────────────────┐
       │ Articles included        │
       │ (n=45)                   │
       └──────────────────────────┘
```
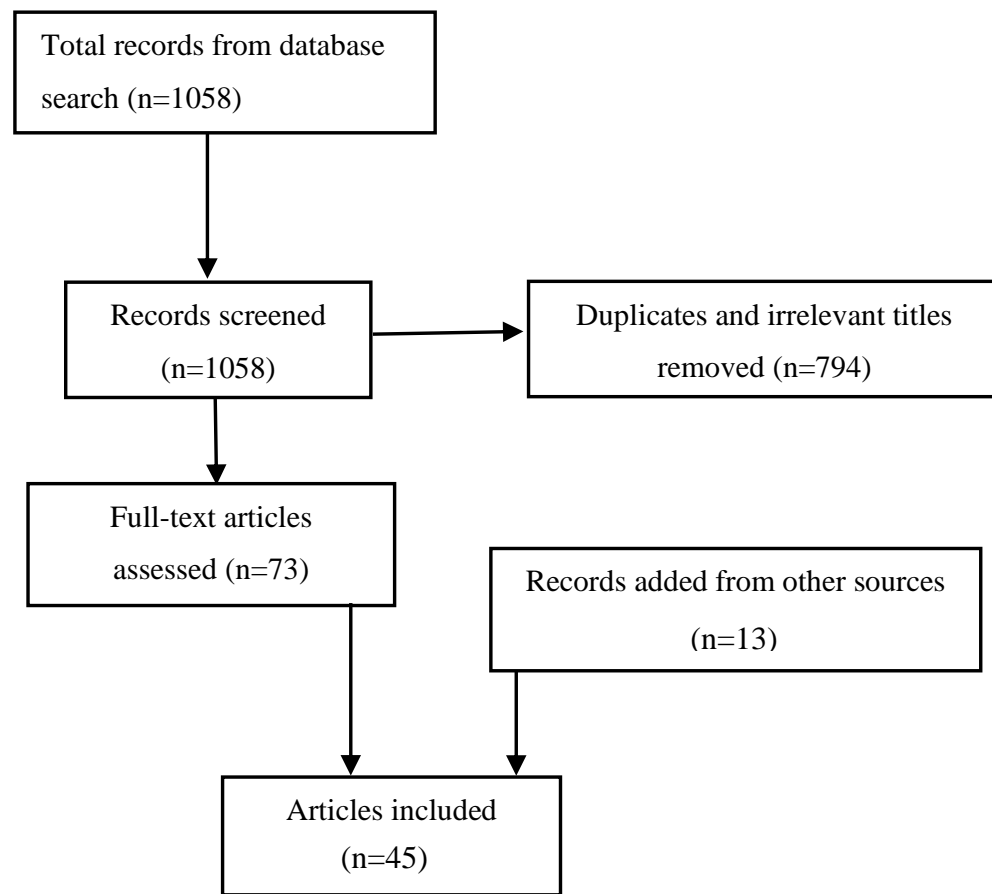
Two reviewers reviewed these and 100% agreement was achieved on their classification. At the same time, 30% of the rejected papers were also reviewed, and a 100% agreement on the reason for exclusion was also reached. After reviewing the reference list, five papers that were not present on the initial search were included: Wallis, et al. (2007); Bruckner, Bartels and Blume (2011); Schmitz, Kroeger and Effenberg (2014); Newbold, Bianchi-Berthouze and Gould (2017); Singh, Bianchi-Berthouze and Williams (2017). On 1st February 2022 an updated search was performed using the existing search strategy 8 additional papers were subsequently identified and included, bringing the total to 13 added from other sources. The final list of papers can be found in appendix 1.

2.3.1   Study type

From the final list of 45 papers, 42 were experimental studies, 21 were randomised controlled trials, 25 were published in full-text peer-review journals, and 15 in conference proceedings; one was an observational study published in conference proceedings.

The results (without accounting with the post-2018 papers) showed an increase in published papers since 2012, with only nine studies published on or before 2011. Overall, 25 studies have focused on motor control in healthy participants, and 13 studies have investigated sonification effects on patients with a medical condition. In these, stroke was the most common condition with four studies (Wallis, et al., 2007; Scholz, et al., 2014; 2015; Singh, et al., 2014), three in low back pain subjects (Singh, Bianchi-Berthouze and Williams, 2017; Singh, et al., 2014; 2016), one with shoulder dysfunction diagnoses (Vogt, et al., 2010), one with Osteoarthritis (Pauletto and Hunt, 2009), one with deafferented individuals (Danna and Velay, 2017) and lastly one with Parkinson's Disease (Gorgas, et al., 2017). The remaining studies, four, used healthy participants but did not disclose the number of individuals (Bruckner, Bartels and Blume, 2011; Brock, et al., 2012; Bruckner, Theimer and Blume, 2014; Jakus, et al., 2017). The total number of experiments reported here is superior to the total number of papers included in the review, as some authors described more than one experiment.

Researchers investigated several activities; a summary can be found in table 2.1.

Table 2. 1 Type of activity sonified

Movement type (total of papers)	Authors
Upper limb movements, such as hitting a target, bi-manual task, writing (15)	Wallis, et al. (2007); Bruckner, Bartels and Blume (2011); Schmitz, Kroeger and Effenberg (2014); Scholz, et al. (2014; 2015; 2016); Vogt, et al. (2010); Danna and Velay (2017); Brock, et al. (2012; 2013); Bruckner, Theimer and Blume (2014); Dailly, et al. (2012); Fujii, Lulic and Chen (2016); Boyer, et al. (2017); Dyer, Stapleton and Rodger (2017)
Trunk movement (4)	Singh, Bianchi-Berthouze and Williams (2017); Singh, et al. (2014; 2016); Newbold, et al. (2016)
Standing balance (4)	Chiari, et al. (2005); Varni, et al. (2007); Giansanti, et al. (2009); Anlauff, Cooperstock and Fung (2013)
Rowing (2)	Effenberg, et al. (2016); Schaffert and Mattes (2016)
Countermovement jump (1)	Effenberg (2005)
Running (1)	Eriksson and Bresin (2010)
Sit-to-stand (1)	Wang, et al. (2014)
Squat (1)	Newbold, Bianchi-Berthouze and Gold (2017)
Golf swing (1)	Jakus, et al. (2017)
Pedalling (1)	Schaffert, et al. (2017)
Ankle movement (1)	Matsubara, et al. (2015)
Walking (1)	Gorgas, et al. (2017)

2.3.2   Data capture

There have been different approaches to capturing and analysing movement, leading to a

number of results; as technology advances, more approaches will appear in the future. In the reviewed papers, more than 13 different types of techniques were applied to analyse human movement. Data to be sonified were captured using a great variety of techniques, as shown in Table 2.2.

Table 2. 2 Technique used for movement capture

Techniques (total of papers)	Reference
IMU (8)	Bruckner, Bartels and Blume (2011); Schmitz, Kroeger and Effenberg (2014); Scholz, et al. (2015; 2016); Bruckner, Theimer and Blume (2014); Vinken, et al. (2013); Chiari, et al. (2005); Giansanti, et al. (2009)
Smartphones with embedded IMU's (7)	Newbold, Bianchi-Berthouze and Gold (2017); Singh, Bianchi-Berthouze and Williams (2017); Singh, et al. (2014; 2016); Bruckner, Theimer and Blume (2014); Newbold, et al. (2016); Eriksson and Bresin (2010)
Optical tracking such as Vicon or Qualysis (5)	Wallis, et al. (2007); Vogt, et al. (2010); Jakus, et al. (2017); Dailly, et al. (2012); Dyer, Stapleton and Rodger (2017)
Microsoft Kinect (3)	Singh, et al. (2016); Anlauff, Cooperstock and Fung (2013); Wang, et al. (2014)
Force plates (3)	Gorgas, et al. (2017); Chiari, et al. (2005); Effenberg (2005)
Grip/Footrest Forces and/or sliding seat movement (3)	Effenberg, et al. (2016); Schaffert and Mattes (2016); Schaffert, et al. (2017)
EMG data (2)	Pauletto and Hunt (2009); Peres, et al. (2017)

Electronic goniometers (2)	Fujii, Lulic and Chen (2016); Matsubara, et al. (2015)
Tablets (2)	Danna and Velay (2017); Boyer, et al. (2017)
USB mouse (1)	Scholz, et al. (2014)
Custom build platform (1)	Varni, et al. (2007)
Breathing sensors (1)	Singh, et al. (2016)
Instrumented insole (1)	Gorgas, et al. (2017)

2.3.3   Effects of Sonification

The effects of sonification were presented in table 2.3. Different sonification approaches either from using sonification as a standalone approach or from a multimodal system. The most common effect was improved motor control and/or movement quality, with ten different papers found. Other effects are found on the next page.

Table 2. 3 Effects of sonification

Effects of sonification (number of papers)	Reference
Improvements in motor control and/or movement quality (10)	Eriksson and Bresin (2010); Gorgas, et al. (2017); Newbold, Bianchi-Berthouze and Gold (2017); Schmitz, Kroeger and Effenberg (2014); Scholz, et al. (2015; 2016); Singh, et al. (2016); Wallis, et al. (2007); Wang, et al. (2014); Schaffert, et al. (2017)
Encouraged movement (7)	Dailly, et al. (2012); Scholz, et al. (2015); Singh, Bianchi-Berthouze and Williams (2017); Singh, et al. (2016); Newbold, Bianchi-Berthouze and Gold (2016); Vogt, et al. (2010)
Improved body awareness in space (2)	Singh, Bianchi-Berthouze and Williams (2017); Varni, et al. (2007)
Increased performance executing complex movements when compared with (no audio) (5)	Boyer, et al. (2017); Danna and Velay (2017); Dyer, Stapleton and Rodger (2017); Scholz, et al. (2015); Schaffert and Mattes (2016)
Increased performance executing complex movements when compared with audio-visual (1)	Effenberg (2016)
Increased performance executing complex movements when compared with non-specific audio-feedback (1)	Dailly, et al. (2012)

Facilitated conveyance of data relevant for physiotherapists and patients (3)	Pauletto and Hunt (2009); Schaffert, et al. (2017); Peres, et al. (2017)
Decrease energy expenditure (2)	Boyer, et al. (2017); Eriksson and Bresin (2010)
Support in postural control (2)	Chiari, et al. (2005); Giansanti, et al. (2009)
Improved range of movement (2)	Vogt, et al. (2010); Wang, et al. (2014)
Increased self-efficacy (1)	Singh, Bianchi-Berthouze and Williams (2017)
Decreased joint pain in stroke patients (1)	Scholz, et al. (2016)
Adequate movement perception from audio information (1)	Vinken, et al. (2013)
Increased perception accuracy of sport-related movements (1)	Effenberg (2005)

Specific issues were found that future researchers should consider: sounds are recommended to be perceived as simple and aesthetically positive (Vogt, et al., 2010), device calibration can be used to increase the subjects' confidence on the set-up (Singh, et al., 2014), and how more frequent audio-feedback (100% feedback compared with 50% feedback) improves retention for an upper limb task (Fujii, Lulic and Chen, 2016). Melodic sonification seems superior to rhythmic sonification, which did not improve learning compared with no-audio (Dyer, Stapleton and Rodger, 2017).

Regarding prototype development, researchers should consider the work already developed by Brock, et al. (2012), Anlauff, Cooperstock and Fung (2013), and Bruckner, Bartels and Hume (2011). Lastly, a note to a review by Schaffert, et al. in early 2019 which focus on a broader perspective in auditory information but that could also contemplate relevant information for researchers in the area.

2.4 Discussion

Based on the results of this scoping review, sonification has the potential to support physiotherapy in several ways. The results demonstrate that the literature regarding physiotherapy and sonification provides significant positive indicators but has limited scope. Several limitations should be taken into account in future research: (1) reduced number of investigations on populations with a specific movement dysfunction; (2) reduced sample size; (3) lack of detail regarding retention of learned movement/task over time when compared with a typical approach. These issues limit the ability to draw definite conclusions on the effectiveness of sonification as an audio-feedback tool for rehabilitation.

The author recommends that high quality, scientifically rigorous studies are needed to investigate the effects and effectiveness of sonification in rehabilitation specific tasks by exploring different sonification mappings in common movement dysfunctions. Less than a third of the studies used patients to investigate the effects of sonification. For sonification to be useful in clinical practice, research should provide a concrete case of effectiveness with specific populations by investigating different intervention areas of physiotherapy, namely cardiorespiratory but also by addressing other specific issues such as falls and injury prevention. Researchers are recommended to use gold-standard

outcome measures in long-term RCTs, hence permitting comparison with typical rehabilitation programs, therefore capturing changes in patient's function throughout intervention as well as data regarding retention of functional improvement with the use of sonification.

In most papers, authors have chosen IMU's and smartphones, likely owing to their ability to capture kinematic and kinetic data in various settings, the fact that they are less expensive and more portable when compared with optical motion capture systems. However, optical motion capture systems are more reliable and accurate, which can have implications on the reported results. It is also interesting to notice the use of Microsoft Kinect as a cheap solution for optical capture, which might facilitate its deployment in health services. Although Kinect is not as accurate as other commonly used optical systems (Pfister, et al., 2014), progress has been made to improve its accuracy (Tanaka, et al., 2018). Researchers should focus on an accessible procedure while improving the accuracy of kinematic and kinetic information collected. At the same time, they should focus on testing different sound attributes for each data parameter in order to optimise effects for functional tasks.

Upper limb tasks were more commonly investigated, with several studies investigating stroke patients and testing manual tasks, with a particular interest in motor control tasks. Current research indicates that sonification appears to facilitate motor learning when compared with other types of feedback (visual) or a sham sonification on complex motor tasks (Scholz, et al., 2015; 2016; Effenberg, et al., 2016). Most effects are reported from small experimental groups with reduced statistical relevance for the results to be extrapolated for clinical practice. However, they are positive indicators for future research (Vogt, et al., 2010; Scholz, et al., 2016).

Sounds attributed to data categories should be assessed in their ability to transmit relevant information for the user. The results show, as Dubus and Bresin (2013) in their systematic review, that this is not common practice. This might have an impact on the effectiveness of the final sonification as other sound attributes might be more relevant for the selected data. Scholz, et al. (2014) are an excellent example of how sonification should be tested in advance, as the authors explored what sound attributes were more effective for an upper limb task before proceeding to their follow-up studies to improve upper limb motor function in stroke patients (Scholz, et al., 2015; 2016). The novel use of sonification for

rehabilitation should encourage authors to explore how different sounds might be attributed to kinetic or kinematic data to maximise treatment effects, which should facilitate the establishment of a sonification framework for rehabilitation.

Results also show that sonification has benefits for an upper limb retraining task (Dailly, et al., 2012; Schmitz, Kroeger and Effenberg, 2014; Scholz, et al., 2015; 2016;) as well as on sports complex movement training (Effenberg, et al., 2016). Sonification seems to positively affect movement perception and action by influencing the mirror neuron system while engaging subcortical structures such as parts of the striato-thalamo-frontal motor loops when the auditory stimulus is congruent with the visualised movement (Schmitz, et al., 2013). Retention regarding the learned task is another critical issue that needs to be addressed if sonification is to be used in rehabilitation; subjects using sonification appear to be able to retain information more easily with increased feedback during an upper limb complex task (Fujii, Lulic and Chen, 2016).

The results that arose from the updated review (since 2018), demonstrate a continuity in terms of research areas, with 3 papers pursuing the investigation on the effects of sonification on stroke patients. In total 4 papers were developed in the area of musculoskeletal physiotherapy and two papers focused on sports related movements including static biking and golf putting (Maes, Lorenzoni and Six, 2018; O'Brien, et al., 2020). The remaining two papers were performed on the impact of sonification in relation to gait outcomes (Alcaraz, et al., 2018; Reh, et al., 2019).

In the author's view, these questions need to be answered in future research:

What sounds attributed to kinetic and kinematic data aids motor control and facilitates learning for rehabilitation related tasks?

Is sonification use relevant for rehabilitation for musculoskeletal conditions if there is no significant neurological deficit?

How do sonification effects compare with other current techniques on motor control and retention?

The answer to these questions will support the development of technical and theoretical solutions in human movement sonification for physiotherapy.

2.4.1 Strengths and Limitations

This review does not encompass literature that focuses on the technological development of sonification devices but more on the grey area where physiotherapy and sonification merge. This research did not look for papers that refer to the term *audio-feedback* in order to exclude results that were not sonification. However, it is common to find these terms mixed in the literature. These facts, however, might have contributed to the absence of some papers, which might be beneficial researchers in the field. At the same time, studies are heterogeneous considering its population, scope, and objective. Therefore, careful considerations of the reported results should be considered. However, this report conducts the first study of its kind in the area, which can support the growth of sonification in physiotherapy by providing a broad perspective of current research practice while serving as guidance to address existing issues in current practice.

2.5 Conclusion

There are promising results on the use of sonification to support human movement; however, very few studies used reproducible techniques with the potential to be used in a clinical setting. Researchers and technical development teams should focus on solutions that can ultimately serve the practitioner and the patient.

Physiotherapy can benefit from sonification in three different areas, first as a way of analysing data either from human movement or from physiological parameters secondly, as it allows for audio-only data conveyance during a physical assessment and/or intervention. Thirdly, by allowing the development of telerehabilitation tools for assessment and treatment. Promising results have been obtained in neurological rehabilitation and for motor learning, which should encourage further investigation.

3. FRAMEWORK AND DEVICE DEVELOPMENT

3.1 Introduction

This chapter encompasses different subjects, from the strength assessment in physiotherapy, hardware and software concepts for prototype development while exploring how sonification can be used in rehabilitation. The reasoning for the several steps of device development, study design and various intervention decisions are also provided. The author will describe how strength testing is conducted in physiotherapy and the main problems with current knowledge in the following paragraphs. The initial development of the device and its progress through its different versions is also discussed.

Physical examination in physiotherapy encompasses different approaches according to the area of physiotherapy in which one is intervening. Nonetheless, it is safe to say that assessing a subject's capacity to move against resistance – either gravity or additional resistance – is part of most physiotherapists' assessment routines.

Most physiotherapists' assessment of strength using manual muscle tests (MMT) lacks the discriminatory and precise information demanded in a world where accuracy and quantitative measures are paramount in healthcare scenarios. The author will demonstrate why this could be improved and how this thesis is a clear contribution to achieve this. In figure 3.1, the Venn diagram shows how the different areas of knowledge overlap with the central area reflecting the work developed in this thesis.
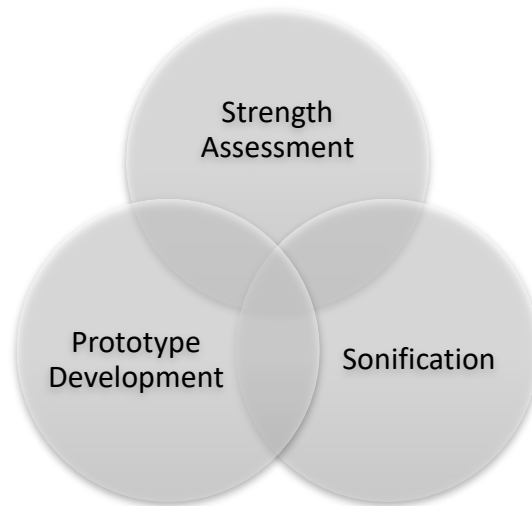
3.2 Muscle function, assessment and task performance

Any type of assessment has a primary goal: to retrieve from the subject, which is the object of an investigation, a measurement that will help us understand its current status. This will allow us to establish that subject's status and compare this with others in similar or distinct situations. Physiotherapists have a crucial role in the rehabilitation process and need to ascertain muscle status - either by group or individually - to devise a rehabilitation plan that can impact individual functional goals. In sum, even though most physiotherapists might want to make the best decision in their clinical intervention, they should be provided with better tools to support their patients.

There are several definitions of muscle strength in the literature, which Enoka (1988) described as a problem for improving research on the theme. At the time, he described strength as "the maximal, voluntary, isometric force"(p.163). Nonetheless, the use of the term isometric is not consensual, and many more definitions can be found in the literature; Sale (1991) describes it as the maximum force of torque produced while performing a maximal voluntary contraction in determined circumstances. Zatsiorsky and Kraemer (2020) defined strength as "the ability to overcome or counteract external resistance by muscular effort"(p.19). Whereas McBride (2016) described it as the "ability to exert force"(p.25), the same author also suggested that researchers disagree on how strength should be measured. This issue arises as different approaches will be measuring other

constructs in terms of strength assessment and can vary from muscle tests to functional tests. An essential idea from Anderson, Madigan and Nussbaum (2007) that supports this affirmation is that the different conception of strength and different possible outputs of muscle strength cannot be adequately expressed by one single value. In this thesis's context, strength is based on the definition of Zatsiorsky and Kramer (2020) and is defined as *the quantitative output from muscle force produced by an active concentric action on a specific joint under external manual resistance and measured by a dynamometer*.

Physiotherapists observe human movement as an integral part of an intricate individual that is influenced by its genetics, environment, and motivation to interact with its surroundings. Movement depends on muscle activity, which will induce a segment or multiple segments to move in relation to a joint. The oldest method of quantifying force is by assessing how much weight one can lift. Physiotherapists perceive muscle strength as a part of their assessment of muscle status and commonly use manual muscle tests to do so.

In general, physiotherapists consider a myriad of approaches when assessing muscle function, more specifically muscle strength, control and length (Petty, 2011). In clinical practice physiotherapists also test strength to: assess patient's progress in acute hamstring injuries (Whiteley, 2018); as part of criteria which can help determine return to play deadlines (Martin, et al., 2022); as part of an injury prevention program (Emery, et al., 2021); to increase task performance (Gary, et al., 2011); and to assess risk of injury (Wollin, et al., 2020).

The use of strength assessment in such varied contexts implies that physiotherapists are aware of physiological *nuances* which can impact results. In practical terms practitioners need to consider several physiological phenomena when testing strength:

a) Length-tension relationship - where the overlapping of actin and myosin filaments dictates the level of performance of muscle contraction, more specifically there is a specific zone of overlapping where the muscle is able to perform at its best whereas excessive overlapping or diminished overlapping between actin and myosin will cause decreased muscle performance (Lieber and Ward, 2011), this is particularly important in isometric testing with regular HHD but also in end or range testing with ASSA.

b) Force-velocity and type of muscle contraction - a slower concentric contraction can elicit a larger force when compared to a contraction at higher speeds, this can be further expanded to other contraction types. With higher muscle force being shown on isometric contractions and even higher with eccentric contractions (Alcazar, et al., 2019). These principles were based on Hill's work from the first half of the 20th century and have clinical implications for research and clinical practice. Specifically, for this research where the ability to resist the movement is paramount for reliable results, slower speeds will be preferred for testing in detriment of higher speeds (the torque produced would be smaller but it would make it difficult for the therapist to resist the movement and maintain the alignment of the dynamometer). The Force-velocity principle is closely correlated with power-velocity (with power being calculated with – force x velocity), in this context, it describes the velocity at which more power is attained for a certain muscle (or muscle group) and has a direct correlation with the capacity to lift a certain load. In practical terms it is represented by a parabolic function where the maximal power output occurs at medium speed and decreases if the speed of the selected movement is either too fast or too slow (Sargeant, 2007).

c) Cross Education - may influence results of muscle output. It has been shown to impact the muscular function of the opposite limb in immobilised individuals and can be used by physiotherapists in clinical practice to maintain strength in specific cases (Mendy, Spittle and Kidgell, 2012). This effect has not been assessed when comparing both limbs in the use of HHD's and might need to be considered as it can affect reliability.

d) Post-activation potentiation - One of the main goals of this work is to be able to elicit and assess the highest strength ability in each participant. However, physiological confounders could hinder the reliability of the device if strength changes occur between repetitions/testers. Such issue might arise from the phenomenon of post-activation performance enhancement, it has been defined as an augmented strength/power output in result of a maximum or near maximum contraction (Seitz and Haff, 2015). A meta-analysis in 2013 concluded that power output was changed by the presence of a conditioning activity and increased effects in individuals with increased training experience but not between genders (Wilson, et al., 2013). The same authors also added that changes were more

evident in multiple sets of exercise, at moderate intensities and with rest between 7-10 minutes. Although some authors have postulated that peak force is not changed (Blazevich and Babault, 2019) another issue still remains; the possible influence of post-activation potentiation might affect the results when performing a manual muscle test between 4 to 8 minutes after the first test if work or angular impulse is being considered.

There are other important terms to clarify which closely relate to muscle assessment when using a dynamometer. They are: 1) torque (defined as the ability of a force to rotate a determined segment around a certain fulcrum) can be seen as an integral system by which movement is described quantitively in joints and can be calculated by moment arm by the applied force; 2) moment arm which can be understood as a lever within the human body usually measured in meters; 3) and finally the moment arising from the gravitational force, which is known to affect the measurement of dynamometers and should be taken into account when calculating force output in these devices (Kellis and Baltzopoulus 1996; Lieber, 2010; McBride, 2016). Further details on how this was performed can be found on section 4.5 of this thesis.

These points can impact the current research and should be considered as confounding factors in future trials. In the next section the authors show how strength can be assessed using several approaches.

3.2.1 Manual muscle tests

The origin of MMT in the United States of America (USA) can be traced to over 110 years ago wherein the Boston Medical and Surgical Journal published an original article by Wilhemine G. Wright (1912) called "Muscle training in the treatment of infantile paralysis" or as it is commonly known now - poliomyelitis. Here she described the work she did as an assistant to a surgeon called Dr Robert W. Lovett using what we now consider MMT principles. In order to face the epidemic at the time, systematic assessment of muscles by doctors was needed. The rise of poliomyelitis outbreaks since the late 19th century led to increased awareness and widespread fear about the disease and its ability to affect mainly children under 16. It was considered a public health crisis, only 10% of those infected with poliomyelitis developed any symptoms, but its effects and cost to

society were abysmal with paralysis (where rehabilitation was needed – hence the importance of strength assessment) and death as a consequence. In 1959, Wintz (1959) defined manual muscle testing as "a procedure for evaluating the function and strength of individual muscles and muscle groups based on the effective performance of a movement in relation to the forces of gravity and manual resistance" (p.466).

MMT is a common type of assessment – Bohannon (2005) points out it is the most common strength assessment method - usually taught to first/second-year physiotherapy students. It consists of a series of movements executed against the therapists' manual resistance in a specific position for both therapist and patient. Manual muscle tests are usually isometric tests (performed in one particular joint position against an ideally immovable resistance) and are known as break tests; but can also be concentric (through the range of motion with muscle shortening) – active resistance tests; or eccentric (through the range of motion with muscle lengthening) – make tests (Hislop and Montgomery, 2007). These tests allow the therapist to understand the user's strength regarding a muscle group/movement. It is a practical skill that requires detailed attention to the patient's position and the physiotherapist's position. This involves training to allow the therapist to use its body weight to counteract the forces exerted by the patient, especially if the lower limb is being tested.

3.2.1.1   *Manual muscle tests scales*

Several scales are used nowadays. Such as the Medical Research Council (MRC) scale or Oxford Scale, graded from 0-5, commonly used by medical doctors, particularly with neuromuscular experience as suggested in a small survey by Dyck, et al. (2005). It is also widely reported in the literature as usually used by physiotherapists, and although no data is available to support this claim, the MRC is traditionally recommended in books for physiotherapists such as Petty, et al. (2011). The MRC scale has been in use since the early 20th century, but others such as the scale from Hislop and Montgomery (2007) or Kendall, et al. (2005) are also used. These two approaches use a 0-5 grading, which also has a correspondence to words such as Normal, Good, Fair, Poor, Trace, or Zero or letters (N, G, F, P, T, 0) - to symbolise their basic grading categories which help establish qualitative information to the test result. Additionally, the scale suggested by Kendall, et

al. (2005) provides the physiotherapist with the added information from plus and minus signs to each of the grades given. Though MMT is still useful for shallow strength levels (for example, neuromuscular problems), users should be aware of its limitations and correct procedures (Palmer and Epler, 1998; Vanpee, et al., 2014; Bohannon, 2018; 2019).

The use of scales with active resistance action has been described as less reliable than isometric tests (Hislop and Montgomery, 2007; Ryder, 2011). However, this has not been thoroughly investigated in the literature using hand-held dynamometers as they are not usually able to gather dynamic data from manual muscle tests. This work aims to analyse the reliability of active resistance tests using a new prototype to detect peak strength during a concentric action.

3.2.1.2   *Practical considerations for manual muscle tests*

MMT using isometric testing, the most used approach in this research field, has shown different degrees of reliability according to the joints tested in systematic reviews (Schrama, et al., 2014; Vanpee, et al., 2014; Chamorro, et al., 2017). Schrama, et al. (2014) demonstrated that elbow flexion and extension displayed acceptable reliability but not for wrist and shoulder movements. This was mainly due to low methodological qualities, and further research in this area is needed to establish baseline reliability. To investigate this issue, shoulder abduction reliability is examined in chapter 5. Vanpee, et al. (2014), in their systematic review, attest the importance of assessing strength in Intensive Care Units and found upper and lower limb reliability tests to be valuable and reliable. This may be due to the lower maximal strength levels found in these settings. Chamorro, et al. (2017) investigated absolute reliability and concurrent validity in hip, knee and ankle joints and reported that knee extension and ankle plantar testing were the movements with the lowest reliability. Thus, the authors recommend that these should not be assessed by HHDs but by an IKD. However, these movements have not been compared with IKDs in an active resistance (concentric) test.

More significantly for the concentric MMT approach, these systematic reviews provide a summary of considerations for further HHD tests –to improve reliability and methodological strength - that were taken into account in this approach: body position

and stabilisation should be standardised and provide stability for both tester and device; evaluator and muscle group strength have an impact on reliability and stronger evaluators should perform better in stronger muscle groups – the current approach aims to minimise this by using concentric instead of isometric testing; fixation systems for HHDs should also improve reliability and ought to be considered – positions for testing were chosen with this in mind to limit segment mobility, but no fixation was used due to the nature of the test; individual factors can affect both tester and participant and should be taken into account – activity levels and verbal feedback were standardised to minimise this.

In terms of their statistical approach, Chamorro, et al. (2017) used Intra-class Correlation Coefficients (ICC), which is not recommended to define validity as this type of analysis only assesses correlations and not agreement, limits their findings. More details on this are found in section 6 of the current chapter.

In sum, results from their systematic review show different values of reliability for different joints. There is a variety of different methodological approaches which limit the transferability of the results. Such as the use of various devices, different testing positions and varied statistical analysis. The recommendations from "Guidelines for Reporting Reliability and Agreement Studies" from Kottner, et al. (2011) are followed to minimise the supra cited problems.

3.2.1.3   *Problems with manual muscle tests*

One of the main criticisms of MMT is the lack of granularity. Although Intrarater reliability has proven to be acceptable, performing the technique (MMT) without a hand-held dynamometer reduces the assessment's sensitivity, which induces clinical limitations by reducing the range of available muscular function detail. This was identified over 30 years ago by Wadsworth, et al. (1987) while investigating intrarater reliability when comparing MMT data with and without hand-held dynamometer in five muscle groups with 11 participants. Although their sample was small, they identify early issues with the use of dynamometers in clinical practice that are still relevant today such as the characteristics of the examiners and positions (such as in upper limb) where the tester might find it challenging to stabilise the HHD and provide adequate resistance to the participant.

Over 50 years ago Beasley (1961), reported that a grade 3 muscle group only had 9% of normal force instead of the 50% strength that a grade 3 should correspond. In 1989, researchers also indicated their reservations about the use of MMT in research and recommended the use of dynamometers as an outcome measure instead of MMT (Aitkens, et al., 1989). In line with these findings, Noreau and Vachon (1998) agreed that MMT is not sensitive enough. This was also corroborated by Bohannon (2001), for instance, who argued that although results between an MMT and a hand-held dynamometer in knee extension tests are highly correlated, they have limited ability to discriminate high levels of variance in strong individuals, and an improvement for MMT was in order. A more recent paper Nagatomi, et al. (2016) also revealed their reservation regarding MMT in patients with strength values closer to "normal". They report that in patients with several shoulder conditions, MMT tests could only detect weakness compared to the opposite shoulder if strength was less than 60% of the "normal" shoulder (Nagatomi, et al., 2017). In general, researchers appear to agree that the use of dynamometers should be part of the assessment when using manual muscle tests to minimise the lack of sensitivity of manual muscle tests in patients that can overcome gravity resistance.

Apart from problems with force discrimination, Knepler and Bohannon (1998) also suggest that inter-rater reliability is a problem for MMT, and when possible, tests should be done by only one tester for grades of 3+ and above. In a more recent paper, Fan, et al. (2010) assessed inter-rater reliability in 26 muscle groups in a group of 19 patients and reported high reliability; however, 10 out of 19 of their participants were "simulated", by pretending to have less force than they actually did, which can affect their results. MMTs results can be affected by factors such as the tester's strength, testing position and force direction (Mulroy, et al., 1997; Knepler and Bohannon, 1998; Palmer and Epler, 1998; Schrma, et al., 2014; Vanpee, et al., 2014; Chamorro, et al., 2017).

In general, MMT has shown good correlation with values from HHD (Bohannon, 2001; Perry, et al., 2004; Lee, et al., 2012). However, it is worth pointing out that correlation does not mean agreement and MMT lacks other vital clinical practice factors. For instance, in 2005, Bohannon (2005) shows that important responsiveness information such as minimal detectable change cannot be obtained from MMT and that patients with progress/change in significant strength scores will not be detected by MMT. Due to

aforementioned problems and the increasing body of research around HHD, more and more physiotherapists have been using alternative muscle testing methods. Nonetheless, its use still appears to be more common in research than in clinical practice.

3.2.2   Hand-Held Dynamometer

In 1978, Pearn reported the invention of an HHD by a Frenchman called John Graham-Desaguliers in 1763, whereas the earliest portable dynamometer was developed around 1798 by Regnier (Pearn, 1978). Since then, several models have been available on the market, and current technological advances have made them more widespread, although not to the point of making them an everyday device. Current HHDs are still expensive for most health professionals to use and are not widespread in clinical settings.

HHDs are generally considered easy to use, portable and cheap when compared with IKDs. HHDs can provide a reliable and valid assessment of strength (Stark, et al., 2011). In general, an HHD consists of a load cell that transforms mechanical tension into an electrical signal for accurate measurement (Garcia and Souza, 2020). The use of HHD in isometric testing in specific populations has demonstrated good results, from respiratory conditions (Dowman, et al., 2016) to myotonic dystrophy (Hebert, et al., 2010). As with MMT, the tester resists the movement with the HHD used as an interface between the hand and the segment. Tests usually consist of an isometric resistance in a specific angle, for a certain period, usually 5 seconds. The output from the HHD is varied and can allow for data comparison of force output from torque. Another vital feature of HHDs, when compared to a manual muscle test, is that it is possible to compare clinimetric properties with other approaches as one can calculate, for example, Standard Error of Measurement (SEM), and Minimal Detectable Change (MDC), these points will be approached in the following sections. This, however, does not mean the HHD eliminates all the problems from MMT, as tester strength is still an important factor in the reliability and validity of the results (Wikholm and Bohannon, 1991; Bohannon, 1997).

More recent approaches have used belts or other means of stabilisation to curtail limitations from tester/joint strength to enhance reliability. For example, Toonstra and Mattacola (2013) compared a fixed HHD to an HHD and IKD in knee flexion and knee extension movements and found that test-retest reliability was only acceptable when

comparing the fixed HHD and IKD. Jackson, et al. (2017) also obtained good results with an adapted PVC pipe used as a stabiliser in a small sample of healthy adult runners. In specific populations, the postulated external fixation hypothesis also appears to be superior to a non-fixed alternative: Bui, et al., 2019 demonstrated the increased reliability of a fixed HHD in detecting knee extension strength in patients with Chronic Obstructive Pulmonary Disease, while Nordin, Nyber and Sandber (2020) investigated the use of a new fixated load cell in patients with congenital heart disease. The fixed HHDs are useful and have shown good reliability but are not suitable for all clinical environments and patients. In conclusion, external fixation should be considered for further research and clinical application. Nonetheless, it does not provide a dynamic muscle profile and fails to consider clinical practice situations where stabilisation/fixation might not always be feasible or acceptable.

Information regarding the correlation between strength assessed by a HHD and functional performance has also been investigated before. Data from several previously, peer-reviewed papers have demonstrated the relation between HHD results and physical performance, specifically when considering: Quadriceps strength tested with a HHD in patients with Chronic Obstructive Pulmonary Disease (Rausch-Osthof, et al., 2014); in older adults both community dwelling and living with partial support (Martien, et al., 2014); in patients post anterior cruciate ligament surgery when investigating hop-tests (Sueyoshi, et al., 2017); and specifically hand grip strength which appears to be related to upper and lower body strength as well as functional movements such as sprinting and jumping (Cronin, et al., 2017).

Regarding the limitations from the use of HHD, authors commonly refer to problems in terms of lack protocol for positioning, force initiation, statistical approaches, tester strength and fixation type rather than limitations arising from device design (Chamorro, et al., 2017; Stark, et al., 2011).

In sum, an HHD should be used in clinical practice, if using manual muscle tests, due to considerable evidence in the literature reporting acceptable reliability and validity. It is more objective than MMT and provides increased data granularity, which evidently influences individual muscle strength and exercise prescription accuracy. HHDs have other advantages consistently reported by researchers such as being portable and cheaper

than an IKD which is considered the gold standard (Drouin, et al., 2004; Martin et al., 2006; Stark, et al., 2011).

3.2.3 Isokinetic Dynamometer

IKDs are used comprehensively in research and high-level sports clubs, but they are not easily available. IKDs are considered a gold-standard laboratorial technique due to their reliability and validity (Drouin, et al., 2004; Orri and Dardem, 2008; Habets, et al., 2018). It allows for a standardised assessment with pre-set positions for testing where angular velocity and category of movement action can be easily chosen, which allow for a myriad of outputs to be chosen such as peak torque, work, peak torque-to-body-weight, angle of peak torque and other (Baltzopoulos, et al., 2012; Pescatello, et al., 2014).

Peak torque is one of the most commonly used outputs due to its reliability. With Habets, et al. (2018) reporting, in terms of peak torque for concentric movements, for an IKD (Humac NORM) moderate to good ICC values: for knee extension between 0.76 and 0.86, for Knee flexion between 0.74 and 0.82 while internal shoulder rotation showed ICC values of 0.88 and 0.94, external shoulder rotation had ICC values between 0.81 and 0.94. In terms of SEM(%), the values ranged from 17.1-23.1% for knee extension; 17-22.1% for knee flexion; 9.5-14.4% for shoulder internal rotation; and 6.9-12.1% for shoulder external rotation. Some authors suggest that for anterior cruciate ligament injury around 10% of strength differences are clinically meaningful (Wellsandt, Failla and Snyder-Mackler, 2017); these values might be less than optimal, but they are none the less superior to those from HHDs and MMTs. Peak torque is reported in Newton-meter and when comparing the IKD with the new prototype, results were converted to the same unit (N*m). However, the comparisons between participants using ASSA data were presented as peak force (Kg).

Research regarding angle of peak torque and work is also reported with high reliability levels due to the high correlations reported (Perrin,1986; Perrin, Robertson and Ray, 1987; Montgomery, Douglass and Deuster, 1989). These findings, however, did not clearly ascertain the value of work as a variable from IKDs for future research (Morrisey, 1987). This was later supported by Kannus (1990, 1994) who suggested that work might

not be a better representation of muscle function than peak torque even though it captures performance through the range of motion than one point only.

Nonetheless, a few years later, in 1996, Gleeson and Mercer (1996) quoted several authors indicating that other variables such "standardised angle-specific torque, torques relativised to a dynamic maximal voluntary range of joint motion, and total work" were being investigated and could serve to provide further detail into muscle function. This also seems to be the position of Amaral, et al. (2014), who suggested that total work should be used as an adjunct to peak torque to gather information about the muscle ability to produce torque, as they indicated concerns about the inability of peak torque to be fully discriminative of muscle function through the range of motion.

Recently, with the growing interest in muscle fatigue, more protocols have been devised that explore peak torque, total work and their influence in detecting changes in muscle performance (Bosquet, et al., 2010; Mendonca, et al., 2011; Gautrey, Watson and Mitchel, 2013; Ribeiro, et al., 2015). They seem to be both reliable, but this is still a matter of debate due to the wide range of protocols and approaches to exploring this issue. Work is calculated multiplying force by distance, this is commonly calculated as the area under the curve (torque) versus the displacement of the joint with the SI unit of Joules (Perrin, Robertson and Ray, 1987; Biodex Medical System, 1991; Bosquet, et al. 2015). However, as the author was unsure about the prototype's ability to gather accurate joint displacement it was established that angular impulse could be calculated as an alternative. Therefore, if the data is deemed accurate it could be possible to assess fatigue as an indirect measure of work, this however would need further analysis and development of an appropriate protocol with ASSA. Angular impulse derives from total work as is closely related to work as it also is calculated from the area under the curve, but instead of distance it uses time (Angular Impulse = Torque x Time) (Kreighbaum and Barthels, 1996).

Similarly, recent research regarding range-dependent muscle assessment has questioned the current paradigm focus of peak torque in hamstring risk injury, when it might be more relevant to consider torque changes through the full range of motion (Cohen, et al., 2015; Pieters, et al., 2020). The suggestion to use torque displayed across the ROM are similar to the postulated by Greig and Naylor (2017) when investigating the ability of different ratios (torque based or angle based) to predict performance in an agility test. This makes

it more critical to have tools that a clinician can use to detect torque changes across the available joint range, which this thesis will investigate.

Another commonly recorded parameter from IKDs is the angle of peak torque, it represents the angle of which the highest torque value occurs. It is not as reliable as peak torque with authors reporting values that range from 0.16 to 0.64 (Wilhite, Cohen and Wilhite, 1992) and 0.28-0.67 in knee flexors of a small sample of volleyball players (Dauty and Rochcongar, 2001). The shoulder joint results appear to demonstrate worse reliability than in the knee joint, with variability from 25.1-41.1% (Mayer, et al., 1994). However, other authors reported higher levels of reliability, such as Maffiuletti et al., 2007 in knee flexion (ICC>0.50) and knee extension (ICC>0.90) concentric movements using a Con-Trex IKD and Hill (2014) who demonstrated a moderate ICC for of angle of peak torque ranged between 0.53 – 0.73 in knee extension movements.

With changes in angle of peak torque related to eccentric hamstring training (Clark, et al., 2005); stretching (Cramer, et al., 2007); fatigue (Coratella, et al., 2015) and after anterior cruciate injuries (Eustace, Page and Greig, 2019; Królikowska, et al., 2019) and with physiotherapy interventions able to influence all of the effects above, it makes it attractive for this parameter to be investigated. In recent years, researchers have also used joint angle-specific torque data to obtain the hamstring to quadriceps ratios in order to find possible strength imbalances (Ayala, et al., 2012; El-Ashker, et al., 2017). If the prototype is reliable, this type of data might be able to be retrieved.

Lastly, it is important to refer that testing using IKDs does not reflect a functional movement as it does not allow for speed changes, and most IKDs only allow movement in one plane of movement. Nonetheless, the IKD was used for comparison with the prototype due to its reliability, validity, and considering its aptitude to quantify concentric movements.

3.2.4   Alternative approaches to common hand-held dynamometers

New methods have been developed in the last two decades as an alternative to the common hand-held dynamometers. The dynamometers developed by Li, et al. (2006) and Janssen and Le-Ngoc (2009) aimed to provide a dynamic overview of muscle function at a fraction of the cost of an isokinetic dynamometer by investigating concentric manual

muscle test. These have demonstrated good reliability but have not been followed up in further research or with further prototypes.

New hand-held dynamometers for eccentric tests have also been tested with the work from Karabay, Yesilyaprakand and Picak (2020) where 25 participants were tested for shoulder abduction in order to investigate ICC and validity. ICC was high in their work; however, they used ICC (3,k), which has a tendency to increase the final results. Cadogan, et al. (2011) used a device that could assess the range of motion and strength but could only be used with isometric tests as a regular HHD.

At the same time, low-cost isometric dynamometer alternatives to the currently sold HHDs have started to be investigated in the literature (Oh, Kang and Dvir, 2016; Romero-Franco et al., 2019) and this is of particular importance in developing countries and to physiotherapists around the world that might consider current market devices to be too expensive. To the current project, this means that there is a space for this technology that has not been fully explored and that the overall cost of the prototype should be kept to a minimum in order to be accessible to all physiotherapists.

Previous authors in the area of newly developed dynamometers have, however not approached four essential points that can facilitate the use of a concentric HHD:

1) concentric forces during manual muscle tests are smaller than isometric forces which should facilitate the use of the device when compared with the standard approach
2) the use of concentric forces might make the test easier for less experienced users
3) since the HHDs mentioned previously were developed, the cost of hardware components have dropped which can decrease the cost of its development and facilitate its widespread use
4) sonification has never been tested in HHDs and can provide useful live audio-feedback to improve reliability and patient experience

3.3 Device development

Arduino-based Strength Assessment (ASSA) is a new prototype developed by the author of this thesis which aims to facilitate how physiotherapists and other professionals assess strength in humans. Strength assessment has been up for thoughtful debate since it was

first idealised. However, as previously identified, consistent problems with manual muscle tests signify that it still indicates a suboptimal practice when physiotherapists use manual muscle tests. Some key questions are how to assess strength initially and how to measure its changes throughout treatment. This project aims to change that by using technological advances to support clinicians worldwide with the development of a new hand-held dynamometer.

This portable prototype aims to enhance strength testing using MMT, thus allowing physiotherapists and exercise professionals to perform accurate measures of muscle strength with increased detail about dynamic muscle performance and with a lower price than similar market competitors. The device will provide visual information and audio feedback as an output, which will aid assessment and exercise prescription. The next step of this research will investigate its ability to provide comparable data to a gold-standard technique (validity) and between users (reliability) while assessing its sonification capabilities.

Even though ASSA is considered to be in the same category as HHDs, it provides extra information not commonly found in a standard HHD. HHDs are a common type of device which uses a load cell and a small processor to record muscle strength data from an isometric activity in several different joints. Devices on the market, such as the MicroFET 2 and 3, Lafayette, and others, have embedded microprocessors capable of saving and displaying information or convey it to a computer either by cable or through Bluetooth/wireless connection.

Typical measures include peak torque, time to peak torque, test time, time in a specific torque, and mean torque. ASSA aims to provide dynamic strength and range of movement feedback, similar to previous work by Li, et al. (2006) and Janssen and Le-Ngoc (2009), in contrast with HHDs which only provide static measures of force. In clinical practice and other locations such as health and sports clubs, the development and influx of these devices should disrupt how current manual muscle tests are performed and improve clinical practice. In order to do this, innovation is paramount.

Several versions were developed and went through several rounds of testing to create the prototype's final version, this can be found in figure 3.2. An iterative process was

followed, thus allowing for improvements to the hardware and software used to enhance the device during this project's development.

Figure 3. 2 Prototype development stages

Stage	
Stage 1	•Need highlighed from clinical practice •Literature supports need for developing sonification in physiotherapy
Stage 2	•Investigation of similar products •Creation of proof-of-concept - v1
Stage 3	•v2 Prototype creation •Pilot testing
Stage 4	•v3 Improve prototype •Test prototype in healthy population - Experienced physiotherapists
Stage 5	•v4 Prototype final version •Experiences physiotherapists vs physiotherapy students •Sonification testing
Stage 6	•Focus group
Stage 7	•Regulatory phase development (post-Ph.D.) •Final design and commercialisation (post-Ph.D.)

3.3.1 Version 1

The first attempt to build a dynamometer was made with a load cell from a hand luggage scale (Figure 3.3). The main goal was to establish the idea's feasibility by using simple, cheap components that could be easily manipulated and trialled in an informal environment. Open-source software was downloaded from the internet at this stage, and the hardware was tested in separate to assess the project's viability (Seidle, 2014; Thomsen, 2014).

Figure 3. 3 Hand luggage scale

The first version (Figure 3.4) of the device featured an Arduino Uno® (Appendix 2) as the microprocessor with a low-cost processor. The processor was combined with an MPU-6050 as an inertial measurement unit (IMU) to measure the device's displacement; load cell amplifier HX-711 which transforms data from the load-cell to the processor; and a hand-luggage scale as a load cell to gather data from the force applied. A laptop powered the device. The IMU MPU-6050 (with 16-bit analogue-to-digital conversion) has a three-axis gyroscope and a three-axis accelerometer (range of: +-250º/sec to +-2000º/sec for the gyroscope and +-2G to +-16G for the accelerometer) and requires a power supply of 3-5V (Appendix 3). The load cell amplifier (HX-711) has a 24-bit, analogue to digital converter for weighing scales that need a 5V supply to produce data output at 10Hz (Appendix 4). The hardware was recommended from the tutorials online and are commonly available throughout the world at a low-cost, which is an essential point if the device is to be made available worldwide and cheaply.

Figure 3.4 Schematics from v1
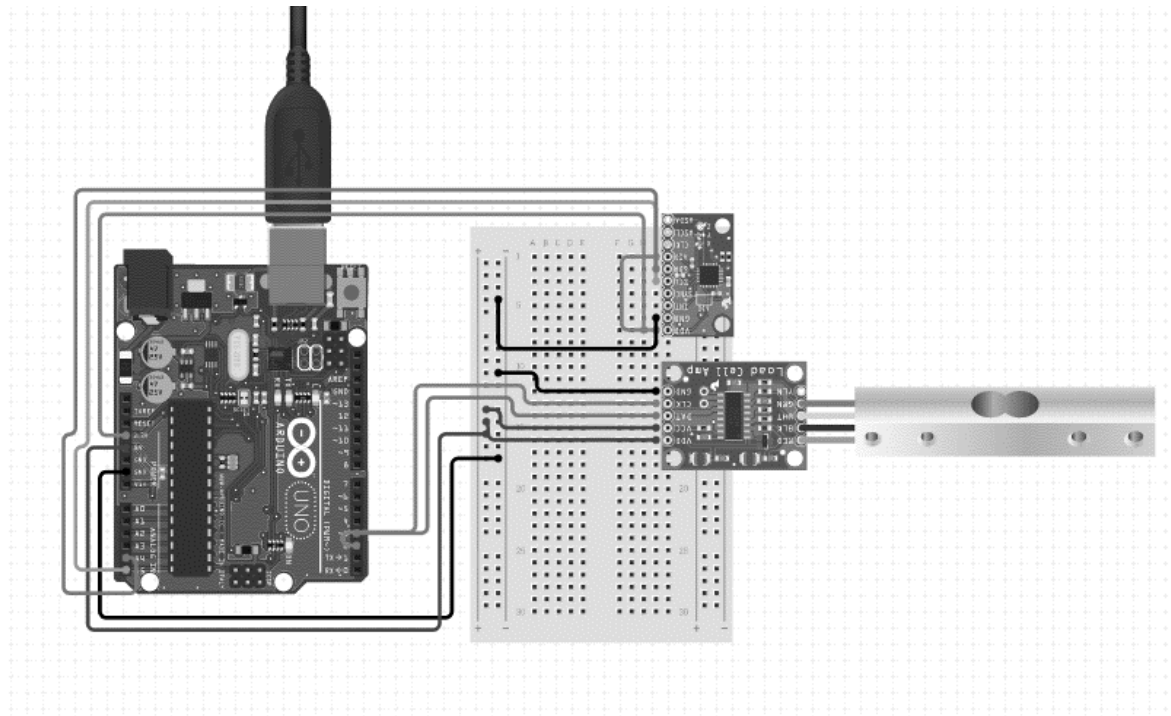
The device showed great promise as it was able to capture information about force (in Kg) applied and movement. Thus, it was decided that further developments were needed to take the project forward. Initial informal contact with several physiotherapists was made to inquiry about the possible use of this device in clinical practice and how it could be developed with clinical relevance. This was the starting point for the initial formal contact with a software developer to support the increased coding complexity needed for the next stage.

The approach that was followed aimed to first establish the project's feasibility by considering the available timeframe, costs, personnel available for the progress, and its final aim. Due to the limited knowledge from the primary author in terms of software and coding, a software developer was approached to build the software. The software developer (VC) was responsible for creating code to intertwine different components and create the user interface. This encompasses the management of each sensor but also the mathematical algorithms to produce accurate results. VC's code and the interface were developed following the author's guidance.

The author used several open-source pages that could compile the data from the different sensors such as but not limited to: For the load cell and load cell amplifier the Arduino HX711 library from Necula (2021); Kalman filter for signal treatment between IMU and processor by Lauszus (2020). At the time of testing it was unknown how the software used could impact reliability as there is no current data on how different algorithms such as the Kalman-filter can impact reliability in this field. It is usually considered that the main limitation arises from the hardware used rather than software (data sheets for hardware used are present in the appendices). The project's primary goal was to provide a solution for physiotherapists worldwide, either by having a low-cost or open software solution that could improve the use of manual muscle tests. Therefore, hardware selection was made with the software developer taking into account several issues such as the programming difficulty, cost, and hardware accessibility.

3.3.2   Version 2

The second version of the prototype (ASSA v2 – Figure 3.5) uses the same Arduino UNO® REV3 as a processor in conjunction with three components: (1) IMU MPU6050; (2) load cell amplifier (HX711); (3) and a new, more accurate, load cell, a PSD S-type Load Cell (575g) with 300kg load capacity (10 x 4 x 4 cm) and 0.02% accuracy (Appendix 4).

It had two wood attachments on each side of the load cell, one for handgrip and one for skin contact. On the side that comes in contact with the patient's skin, two different foam segments improve comfort—one high-density foam with 5mm height and a low-density foam with 40mm height.

The load cell was chosen considering its cost and the mean load that these devices usually undergo, as seen in previous research. As this was a prototype, it was decided to use a cheap S-type load cell with 300kg maximum load (although load capacity is higher than usually required even for lower limbs). The device's output could be seen in real-time on the laptop and was provided in angle and kg (active resistance mode) or kg (isometric mode). The load cell amplifier used was an HX711 with 24-Bit Analog-to-Digital Converter for Weigh Scales to output data at a rate of 10Hz. Finally, Arduino was connected to a laptop through a USB AB cable (2m). The test results were given in

maximum kg force exerted on a determined angle; this means the software filtered data output per angle of movement analysed while minimising unwanted movement by the therapist. Output for angle and force was provided on an Excel® spreadsheet for *post hoc* analysis.

Figure 3.5 ASSA prototype v2- lateral and superior view
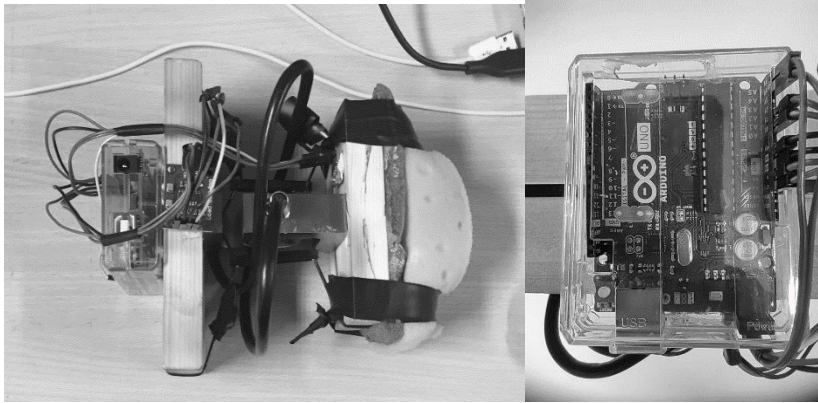
3.3.3 Version 3

Version 3 used the base hardware from v2 except for a new smaller load cell – beam type. This load cell weighted 200g, which was a reduction in weight of over 60%, with the main aim of reducing the total prototype size and facilitate its handling with one hand only.

Figure 3.6 Beam load cell
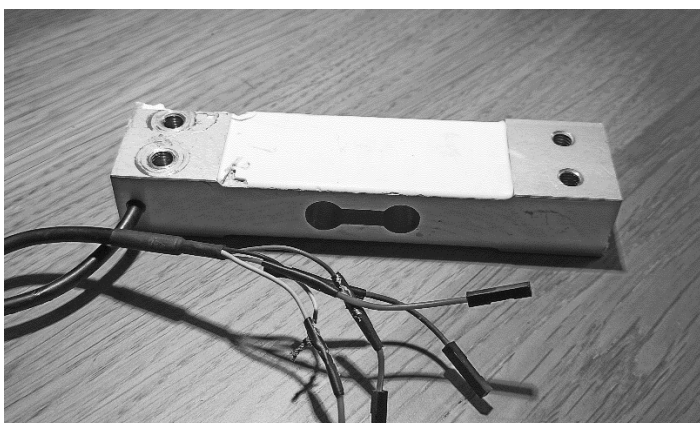
A new outer casing (15cm x 5cm x 8cm) (Figure 7) was used to allow for a smaller and lighter device – figure 3.6.

Figure 3.7 New electronics box
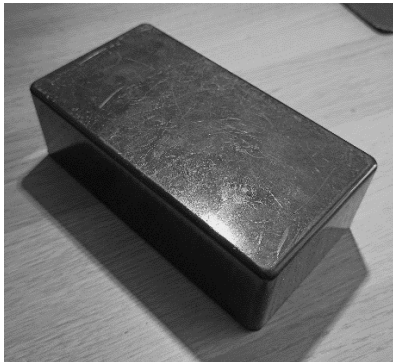
This meant opening the electronics box (figure 3.7) to create a new attachment to the load cell that would serve as the contact area with the segment to be tested. This attachment was made of a wood square (9cm x 9cm x 1cm) with a high-density foam – 1cm height (x2) and a total thickness of 2cm. The final prototype v3 has a total height of 10.5cm, a length of 15cm, a width of 8cm, and weighed 869g (Figure 3.8).

Figure 3.8 Prototype v3 – lateral and superior view
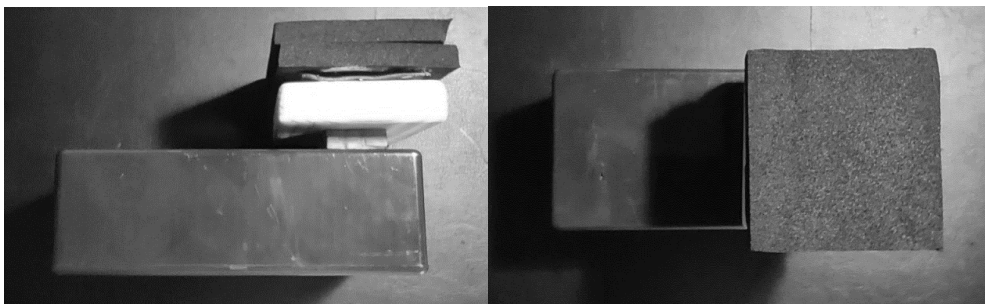
3.3.4 Version 4

Version 4 maintained the hardware from version 3, but two new features were introduced: speed measurement and sonification. The speed output was created as an addition to the output (spreadsheet) as this type of information was already available from the hardware

but only not in use. To create the sonification feature of the device, PyGame (Shinners, 2011) was used. PyGame is a free tool commonly used in game design with Python, which meant a seamless integration with the current program that can be embedded in future devices. This is an innovative alternative to other approaches when using sonification that usually uses an external interface for sonification. It also meant that no added costs were needed to add this feature. In this prototype, PyGame was used as the interface between the data collection process and the sound output (Figure 3.9).

Figure 3.9 Interaction between ASSA and PyGame
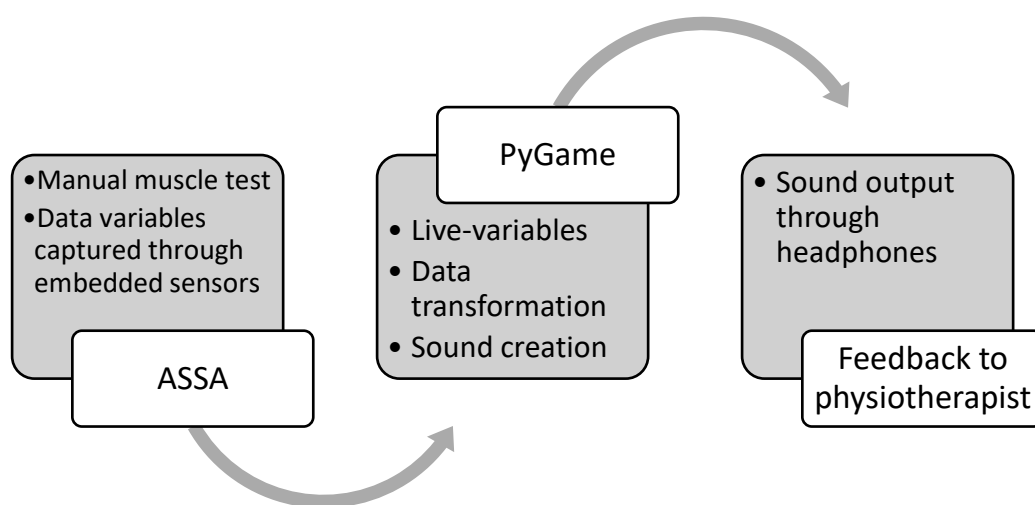
Data variables were captured throughout the range of motion while the physiotherapist performed the movement. PyGame relies on that data stream to create sound - data sonification. Further details regarding sonification are presented in the next section.

3.4   Sonification

3.4.1   Background

The study of sound waves dates to Pythagoras and Aristotle, the latter started to explore the concept of sound propagating in air and that notes travel at different speeds. Simultaneously, its evolution can be traced in the last millennia from Galileo to Marin Mersenne, called by many the father of acoustics, who lived between 1558-1648 and investigated the speed of sound in the air. Another essential researcher was Robert Boyle, who pioneered the scientific method and contributed to the knowledge of the need for

sound to have a medium in which to propagate. More recently, in the early 19th century, Fourier established the Fourier transform, which provides a tool to analyse a sound wave representing time and frequency (Bohn, 1988). While in 1857, the first phonautograph was created. It was used to record sounds and was invented by Édouard-Léon Scott de Martinville - later Graham Bell worked on developing the phonautograph (which Thomas Edison also did) before patenting the telephone (Feaster, 2010).

More recently, with the upsurge of technology, sonification has facilitated human-computer interaction and provided an alternative insight from data. Sonification has been employed in several domains, such as the use of earcons (the sound an icon makes when we click on it), or for data analysis to investigate multivariate streams of information such as salmon migration (Hegg, et al., 2018), to explore spacial data from the Kepler telescope (Winton, et al., 2012) and to aid image-based medical diagnosis (Gionfrida and Roginska, 2017). This highlights how sound and sonification can be used in relevant ways to explore natural phenomena. Physiotherapists and other health professionals, specifically if they work in a hospital setting, are flooded with sounds almost every hour of their workday, either by an alarm or audio display from devices used in clinical practice. However, one cannot forget how useful and practical these sounds are, predominantly saving time and facilitating health care assistance. For physiotherapists, sonification can be adopted using data-related audio-feedback to support human perception in assessment and rehabilitation for both patient and physiotherapist.

Physiotherapy practice has changed considerably in the last 30 years, and that change has can also be associated with the increase in the percentage of physiotherapy related research published within the domain of human-based health (Jesus, et al., 2020). Simultaneously, the amount of research and innovative approaches to rehabilitation has grown, with relevant work arising from the overlap of physiotherapy and sonification (Guerra, et al., 2020). Sonification has the potential to improve the patient's related outcomes due to its specific and systematic approach. Kramer, et al. (1998) defined it as "the use of non-speech audio to convey information". In their words, "it is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation". In this work, sonification will convey live audio feedback to the user, but this can be expanded to provide feedback to the patient as well.

While auditory display is a broad term for sound representations on information from a computer to the user, sonification is more specific. Sonification excludes speech-like audio through the use of audio feedback and has been widely tried in different circumstances (Dubus and Bresin, 2013). Physiotherapists are now, more than ever, using these new technologies in everyday clinical practice. Simultaneously, patients are beginning to embrace such technologies with the added benefit of increasing patient satisfaction. One example is the work developed by Kairy, et al. (2013), where a telerehabilitation system was used for patients after a total knee arthroplasty.

The reason why sonification is different from regular audio-feedback systems can be found in this description by Hermann (2008) "a technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering)". While adding, it should be objective regarding the input data, being systematic and reproducible (Hermann, 2008). This is where sonification differs from common audio-feedback as the latter does not need to have any of these characteristics.

Sound is concurrent with human existence and human activity. From the sound of a beating heart to the inherent familiar sound of walking, these sounds help us make sense of our surroundings, the disposition of other humans and are part of our daily lives in more contexts than we usually perceive. Research from different fields such as cognitive neurosciences, music therapy, and sports has investigated how sound affects our ability to understand and produce movement (Scholz, et al., 2016; Danna and Velay, 2017; Maes, Lorenzoni and Six, 2019). Physiotherapy has a close connection with movement and function, and sonification has shown the potential to assist in human movement perception (Schaffert, et al., 2019). The importance of sonification and audio feedback has been increasing within motor control research with plenty of room for expanding their use to aid rehabilitation. Therefore, and considering the current limitations in the use of manual muscle tests, it is relevant to assess, for the first time, how sonification can impact clinical practice in this specific task.

Schaffert, et al. (2019) provided an overview of the relation between sound and movement in sports and rehabilitation. Their work encourages evidence on the use of audio feedback to support motor learning and sports, revealing its essential role in auditory perception in several tasks with relevance for performance and rehabilitation. Dubus and Bresin (2013), argue that the constant presence of musical characteristics charted to different dynamic

or kinematic movement parameters during the execution of specific movements contributes to "improve movement quality and motor (re)learning" through various parallel pathways which improve performance through error minimisation and improved internal movement representation. All of which can reduce the influence of perceptual impairments. In theory, this increment in performance by minimising error in a specific task should be transposable to different areas within rehabilitation, namely assessment techniques that rely on the systematic performance of a particular gesture. This is the starting point for the exploration of sonification to improve MMT.

Transposing data parameters to sound that the user can easily perceive is the primary goal of using sonification, even though this is not an automatic process but a planned intervention where data source and sound output have to be considered to warrant a clear understanding of the underlying information. If this is not correctly done, the goal of sonification is lost. One of the initial steps of building a stimulating and relevant sound output from human movement data begins by choosing the interface that will process data. In this case, it was the Arduino Uno, as previously explained. To transform data into sound, the software PyGame was used, which acted like a processing tunnel between data input (Arduino) and sound output to the physiotherapist by using Musical Instrument Digital Interface (MIDI) signals.

3.4.2   Sonification for manual muscle tests

According to Huber (2012) MIDI is a language that allows communication between different instruments, hardware and controllers in a particular network. The technology was created to allow for a standardised protocol that could be used by different manufacturers and in different hardware devices in a synchronised manner, which was a problem in sound design before the appearance of MIDI.

In the case of ASSA, data is collected from the Arduino from several sensors with each data parameter attributed to a different sound parameter. The Arduino output values are then charted to MIDI notes on PyGame, with each value varying from 0-127 or 1-127. The parameters to be modelled are Rhythm, Note, Instrument and Volume. Although PyGame was not explicitly designed for sonification purposes, it can still be used to

connect the input from the Arduino processed data with the live output data from the IMU and load cell in a meaningful form.

Creating the code for the sound feedback implies generating a series of commands that can be understood by the processor and can result in an audible output for the user. In order to do that, it is necessary to understand the basic MIDI commands that will convey audio-specific commands to the controller. In this case, each MIDI value corresponds to a determined frequency (Figure 3.10).

Figure 3.10 – MIDI frequency table

MIDI number	Note name	Keyboard	Frequency Hz	
21 22	A0		27.500	29.135
23	B0		30.868	
24	C1		32.703	
26 25	D1		36.708	34.648
28 27	E1		41.203	38.891
29	F1		43.654	
31 30	G1		48.999	46.249
33 32	A1		55.000	51.913
35 34	B1		61.735	58.270
36	C2		65.406	
38 37	D2		73.416	69.296
40 39	E2		82.407	77.782
41	F2		87.307	
43 42	G2		97.999	92.499
45 44	A2		110.00	103.83
47 46	B2		123.47	116.54
48	C3		130.81	
50 49	D3		146.83	138.59
52 51	E3		164.81	155.56
53	F3		174.61	
55 54	G3		196.00	185.00
57 56	A3		220.00	207.65
59 58	B3		246.94	233.08
60	**C4**		**261.63**	
62 61	D4		293.67	277.18
64 63	E4		329.63	311.13
65	F4		349.23	
67 66	G4		392.00	369.99
69 68	**A4**		**440.00**	415.30
71 70	B4		493.88	466.16
72	C5		523.25	
74 73	D5		587.33	554.37
76 75	E5		659.26	622.25
77	F5		698.46	
79 78	G5		783.99	739.99
81 80	A5		880.00	830.61
83 82	B5		987.77	932.33
84	C6		1046.5	
86 85	D6		1174.7	1108.7
88 87	E6		1318.5	1244.5
89	F6		1396.9	
91 90	G6		1568.0	1480.0
93 92	A6		1760.0	1661.2
95 94	B6		1975.5	1864.7
96	C7		2093.0	
98 97	D7		2349.3	2217.5
100 99	E7		2637.0	2489.0
101	F7		2793.0	
103 102	G7		3136.0	2960.0
105 104	A7		3520.0	3322.4
107 106	B7		3951.1	3729.3
108	C8	J. Wolfe, UNSW	4186.0	

Source: *Wolfe, 1997*

Another perspective into the MIDI values is displayed on Figure 3.11, where octave numbers are also considered - these features allow even an untrained user to create and manipulate MIDI in a simplistic manner as this can be easily changed within the software.

Figure 3.11MIDI values by octave

Octave number	Notes											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	0	1	2	3	4	5	6	7	8	9	10	11
1	12	13	14	15	16	17	18	19	20	21	22	23
2	24	25	26	27	28	29	30	31	32	33	34	35
3	36	37	38	39	40	41	42	43	44	45	46	47
4	48	49	50	51	52	53	54	55	56	57	58	59
5	60	61	62	63	64	65	66	67	68	69	70	71
6	72	73	74	75	76	77	78	79	80	81	82	83
7	84	85	86	87	88	89	90	91	92	93	94	95
8	96	97	98	99	100	101	102	103	104	105	106	107
9	108	109	110	111	112	113	114	115	116	117	118	119
10	120	121	122	123	124	125	126	127				

Source: *Schuyler and Eun-jin, 2016*

Previous research and the task's specificity were considered to select the more meaningful variables for sonification. A few possibilities within the domain of sonification allow for data to be harnessed with an auditory goal. A few examples are audification, parameter sonification, model-based sonification, event-based sonification or earcons. The main feature of the use of sonification in this system is its ability for the user to interact with it. This allows for sound creation which has a direct implication on the type of sonification selected.

By referring to the "sonification space" presented by Ludovico and Presti (2015), one can place the current project within a specific area of sonification. The "sonification space" uses two different axes (x-axis – time granularity and y-axis – sound abstraction level) to suggest a distribution of sonification approaches according to these two variables. According to the data available for this project (from the prototype) and the main goal (clear instructions for the physiotherapist given by interaction with the device and participant), the time granularity available to us should be regular to allow for live reaction to the sound received. In contrast, the sound output should be closer to the graph's low-level area as the task is of short duration, and simpler sounds facilitate the comprehension of the data meaning. This choice is also based on the fact that different groups of people more widely understand low-level sounds than high-level sounds (such as soundscapes), which might differ from culture to culture (Ludovico and Presti, 2015). In terms of the time granularity, it is closely related to the device's data output, in this case at least 10Hz, thus providing a stable output and not a continuous one, making a case for a model-based sonification. Regular time granularity, rather than continuous, provides

more benefits for interactive sonification, which was the case. This fact positions the current work either in parameter mapping or model-based sonification. However, considering that user and participant interaction is the most crucial feature, the current approach can be demarcated as model-based sonification.

Model-based sonification can be defined "as the general term for all concrete sonification techniques that make use of dynamic models which mathematically describe the evolution of a system in time, parameterise and configure them during initialisation with the available data and offer interaction/excitation modes to the user as the interface to actively query sonic responses which depend systematically upon the temporal evolution model" (Hermann, 2011). Model-based sonification was derived from sonification mapping, with the main difference is that it needs energy from the system it refers to create sound, whereas a sonification mapping does not need input for sound to arise (Hunt, Hermann and Pauletto, 2004). In designing the sound models for the current approach for manual muscle tests, the initial primary focus was to provide an increased perception of the interaction between the therapist and the participant, either from the force, speed or acceleration of the tested segment. Once again, this points to a model-based sonification. As noted by Hermann (2011), a few points differ between model-based sonification and parameter mapping sonification. Namely, for model-based sonification: the interaction between the model and the participant is a feature, fewer parameters are needed, there is no need for the creation of "musical listening" as sounds can be more straightforward and the parameter attribution does not need to change for each dataset as in parameter mapping sonification.

Different sonification approaches might help different users in different situations. For example, an experienced physiotherapist using ASSA might benefit from having segmental displacement sound-feedback on a strong patient, whereas force feedback might be more relevant to a novice user. In the future, the prototype could also provide parallel information to the user (patient or client) while focusing on a particular exercise. Sound feedback is of particular importance to provide temporal precision and reduce variability on the task (Kenyon and Thaut, 2003). Most uses of biomechanics data and its effects on motor learning are focused on the feedback to the participant (Effenberg, 2005; Schaffert, Mattes and Effenberg, 2011, 2017; Sigrist, et al., 2013), but this thesis focuses on audio feedback to the physiotherapist and not the patient/client.

In this project, different models were tested to investigate how different sounds from several parameters could provide valuable and understandable feedback on muscle testing in different joints. Dubus and Bresin (2013) published a systematic review that provided an excellent insight into models researched in several sonification areas. The authors provided an overview of areas of sonification regarding the use of sound parameters and have identified several categories that commonly arose from the literature: pitch-related; timbral; loudness-related; spatial and temporal. They have found that the most used auditory dimensions are pitch, loudness and duration, and these are standard features in research using kinetic and kinematics data sources. This work helped the author to select these three parameters (pitch, loudness and duration) as variables to be manipulated. It is important to refer that Dubus and Bresin (2013) state that the commonly used auditory attributes do not automatically make them ideal for every area and that researchers should explore several options according to their goals. However, considering this was the first time sonification was used for manual muscle tests, it made sense to explore these parameters first.

The developed software allows the therapist to manipulate the instrument, pitch, note length, volume, and relationship with manual muscle test variables. The variables selected to manipulate are quite restricted when considering the complexity of professionally created sound design and the endless possibilities of musical creation. Nonetheless, in this work, the sound is used to support and reduce the task's effort rather than embellishing it unnecessarily with layers of sound feedback.
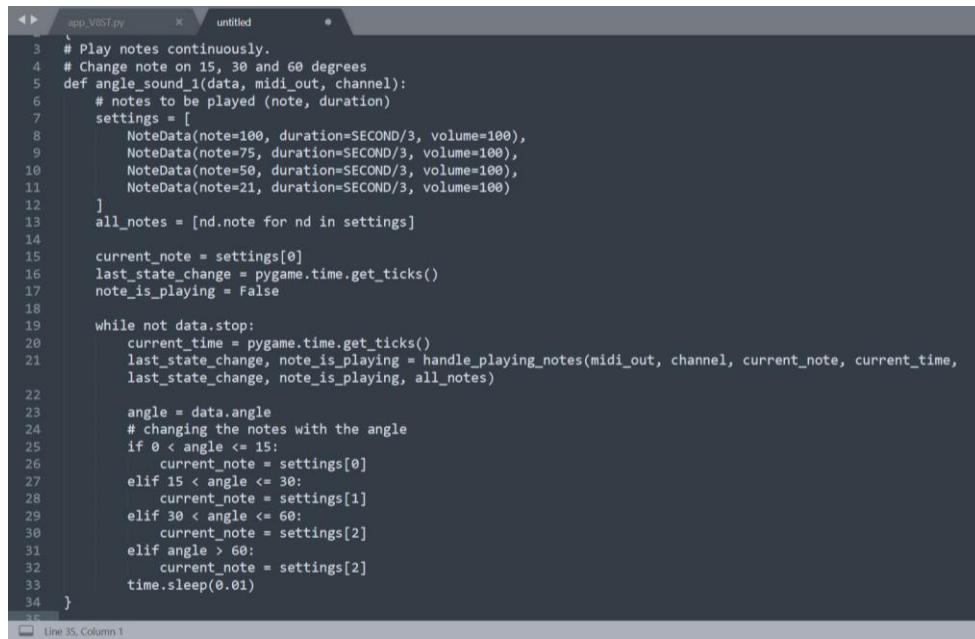
The data parameters from the device and the MMT that can be considered for sonification are time, force, angle, speed and acceleration. Below is an example of the possibilities of PyGame as a tool for sonification used within the context of manual muscle test. This code segment describes PyGame being initiated and the user's ability to select, on the source code, which instrument correspond to each variable (Figure 3.12). In this case, three instruments can be selected for the model-based sonification (line 21-23 in figure 3.12), even though only one was chosen to avoid auditive overload. The list of instruments is readily available in speciality books and websites. The current model used an acoustic grand piano due to its neutral and pleasant tone.

Figure 3.12 Code for PyGame initiation and instrument selection

```
app_V8ST.py          ×        untitled        ●
1
2
3    {
4    def setup_pygame_audio():
5        if pygame.get_sdl_version()[0] == 2:
6            pygame.mixer.pre_init(44100, 32, 2, 1024)
7
8        pygame.init()
9        screen = pygame.display.set_mode((400, 300))
10
11        if pygame.mixer and not pygame.mixer.get_init():
12            print('Warning, no sound')
13            pygame.mixer = None
14
15        pygame.midi.init()
16
17        port = pygame.midi.get_default_output_id()
18        print ("using output_id :%s:" % port)
19        midi_out = pygame.midi.Output(port, 0)
20
21        midi_out.set_instrument(0, channel=0)    # angle instrument
22        midi_out.set_instrument(0, channel=1)    # force instrument
23        midi_out.set_instrument(100, channel=2)   # velocity instrument
24
25        return midi_out
26   }
27
```

The code segment present in figure 3.13 focuses on the charting of sound variables to the speed of the segment moving and how information about speed can be attributed to specific angle ranges for a particular join, it is divided into two parts. Firstly it allows the researcher to choose, under the variable "NoteData", a particular pitch (denoted by "note"), a specific length for that sound ("duration") and an exact volume ("volume") – together, these variables create a "setting" which will then be attributed to the property of interest, in this case, joint angle. Secondly, the researcher or physiotherapist can chart each setting to a different angle range, resulting in different sounds as the angle varies across the test.

Figure 3.13 Pygame Code for selection of parameters

```
app_VBST.py          x    untitled

 3  # Play notes continuously.
 4  # Change note on 15, 30 and 60 degrees
 5  def angle_sound_1(data, midi_out, channel):
 6      # notes to be played (note, duration)
 7      settings = [
 8          NoteData(note=100, duration=SECOND/3, volume=100),
 9          NoteData(note=75, duration=SECOND/3, volume=100),
10          NoteData(note=50, duration=SECOND/3, volume=100),
11          NoteData(note=21, duration=SECOND/3, volume=100)
12      ]
13      all_notes = [nd.note for nd in settings]
14
15      current_note = settings[0]
16      last_state_change = pygame.time.get_ticks()
17      note_is_playing = False
18
19      while not data.stop:
20          current_time = pygame.time.get_ticks()
21          last_state_change, note_is_playing = handle_playing_notes(midi_out, channel, current_note, current_time,
22          last_state_change, note_is_playing, all_notes)
23          angle = data.angle
24          # changing the notes with the angle
25          if 0 < angle <= 15:
26              current_note = settings[0]
27          elif 15 < angle <= 30:
28              current_note = settings[1]
29          elif 30 < angle <= 60:
30              current_note = settings[2]
31          elif angle > 60:
32              current_note = settings[2]
33          time.sleep(0.01)
34  }

Line 35, Column 1
```

The model-based sonifications were developed to provide relevant feedback with the final user in mind, not necessarily to create a pleasant aesthetic sound (Dubus and Bresin, 2013). It is also recommended that the sound-feedback be appropriate for the environment in which the task will be performed - e.g.: user cannot listen to the sound due to other auditive environmental inputs (Walker and Nees, 2011).

This process is not trivial, particularly in sonification applications, as research suggests that an overload of auditory information has detrimental effects on task performance (e.g., Wolf, et al., 2011) and that task-irrelevant auditory stimuli are strong distractors (Parmentier, 2014). The use of meaningful auditory information is, therefore, determinant for the user's experience (Effenberg, et al., 2016; Dyer, et al., 2017a) and needs to be considered in a clear framework for model-based sonification derived from a better understanding of the processes underlying motor learning/control from a basic research perspective (Dyer, Stapleton and Rodger, 2015).

The previous points support the case for a sonification that can be easily understood even by an untrained user. Therefore, the current model-based sonification aims to provide simple and clear audio-feedback instructions to therapists to provide complex information about kinematic data in a clinical setting. To create a sonification model without considering the final auditive objective might create either an innovative display or a complete disaster. Which is why it is of considerable importance that researchers create

a sound design with the end-user in mind. To achieve this, sound feedback was tested in simulated tasks in an environment similar to what physiotherapists might regularly face in terms of noise and activity. Walker and Nees (2011) suggest that when designing sonification three main points should be considered: data-to-display mapping (which data dimension correlates to each sound dimension); the polarity of the map (is the sound attribute chosen positively or negatively related to the data attribute, e.g., positive polarity – increase in force mapped to increase in pitch); scaling of the data attributes (how much sound variation corresponds to each data point). Although these points were created for parameter-mapped sonification, they are widely accepted in most sonification designs. They were also integrated into the process to establish the model-based sonifications for elbow flexion and knee flexion- the joints to be tested.

3.4.3   From data to sound

A six-step framework for the creation of model-based sonification design was suggested by Hermann, 2011. Below is the outline of that framework and how it was used to develop the sonification design.

1 - **Model set up** - In establishing a successful sonification for human movement, or other, the first step consists of determining the data which will be used. In this current prototype, it was muscle force, angle and angular velocity were the available data output. However, only force and angular velocity data were considered to improve reliability at this stage. Data input will differ for each participant, but the same data parameters will be used in the sonification design.

2 – **Model dynamics** - The next step is to identify which sound dimension will correspond to each data parameter. This is of particular importance as it must allow the user to clearly understand the analysed variables (Walker and Nees, 2011). Although this is true for sonification in general, this is quite relevant in this research, as it is hypothesised that the expected effect can provide significant input to the user so that it has considerable effect in the results.

Two main maps were developed one corresponding to force – more force applied meant a higher pitch (data were scaled separately for elbow and knee movements (force data); and another one for segment movement (angular velocity data). In terms of the sounds

68

used, there was a need to maximise the sound range available due to the high number of strength outputs. Therefore, frequencies varied from 27.5Hz to 4186Hz, similar to a grand piano – this was alike for all sound models. Maximal note duration was set to $1/10^{th}$ second to minimise disruption of perceptual interpretation and to maximise auditive pleasantness.

In terms of the data to sound mapping, a linear mapping approach was followed with the lowest end of the frequency mapped to the lowest range of force values from 2kg of force onwards. In turn, this allowed the system to remain silent until the initial force was applied. The end of the frequency range (4184Hz) was attributed to each joint's highest values of force. Pitch was selected for the model-based sonification as it has been extensively used in previous research with kinematics related approaches while being an easily recognisable feature of sound (Dubus and Bresin, 2013)

3 – **Model excitation** – In order for the participant to interact with the system, two different steps were needed. Firstly, the tester would position the device, prepared to start the manual muscle test, applying the correct positioning of the device in the contact area with the segment to be tested and a 90º orientation with that same segment. Secondly, the main tester would then inform the support tester to initialise the program on the computer. Once this was done, the data started to flow, and the interaction between participant and sonification would start.

4 – **Initial state** – If no movement or pressure was present as the program started then no sound would be conveyed to the tester, but data would be collected, nonetheless. This set-up was deliberately designed into the program so the tester would only start the test once ready.

5– **Model link variables** – This step is directly related to how different data variables can become connected to different sound parameters. However, in this case, data features only match one sound parameter due to the short length of the task and eventual latency issues.

6 – **Listener characteristics** -The central aspect is to determine how the user will be receiving the sound (location, orientation or distance between sources), which can be divided into macroscopic or microscopic models. Macroscopic models are more complex and can include soundscapes and audio-panning, whereas microscopic models are single-

sourced and simpler. As the physiotherapist performs a physical task - manual muscle test - that requires attention and effort, consequently, the microscopic model was chosen - with audio being provided by earphones.

The scaling of the data parameters is of particular importance for data collections with a considerable amount of data entries and where the user needs to identify small but significant changes from the sound output. For the current project, this is relevant as the participant's changes in force, speed, and angle need to be identified by the user, and adequate scaling allows the information to be clearly understood. Scaling is a particularly problematic point in designing sonification for manual muscle test, as this has never been done before, and there is no previous research that could indicate a potential direction or any major potential problems.

Lastly, it is crucial to understand who is on the receiving end of the sonification. Physiotherapists are not acquainted with this type of approach; therefore, the user's inexperience can significantly impact how complex the model might be. Furthermore, if physiotherapists were to use this in clinical practice, not much time could be spent learning which sound parameter corresponds to each data variable. Therefore, sound feedback design needs to be clear and conspicuous. Further details from sonification and the full code for the specific use case in manual muscle tests can be found in Chapter 6.

3.5   Measurement tools and properties

In order to use a measurement tool, clinicians and researchers need to achieve a certain level of performance before it can be deemed appropriate and accurate for clinical practice. To achieve that, it is essential to define what a measurement is in the first place. Hopkins and Stanley (1981) defined measurements broadly as "rules for assigning numbers to objects in such a way as to represent quantities of attributes". While Carter, et al, (2011) defined it as "the systematic process by which things are differentiated", the authors add that their definition underlines the fact that measurements need to have specific rules and guidelines as part of their definition process.

It is reasonable to argue that the only reason to measure something is to obtain accurate and factual information about the measured subject. However, there is no perfect

measurement tool, any device ever created by man has some inherent degree of error, even if small. There is no point in collecting data if the device's error is so large that it will remove significance from the analysed information. For instance, if we accept an error of 1° in a cooking thermometer, that might be acceptable, but it would probably not be satisfactory for measuring a chemical reaction where differences larger than 0.01° might change the end product for a particular medication. To detect changes in a certain variable, one must define what is being measured first.

In ASSA's case, the construct to be measured can be defined by the objective quantitative output from peak muscle force produced by a concentric action on a specific joint under external manual resistance and measured by a dynamometer. In the context of this thesis, it is also crucial to state that muscle force is represented by weight (in Kg) or torque (in Nm) both considered ratio scales to be measured as continuous variables (Carter, et al., 2011).

It is particularly relevant for researchers to determine what will be assessed when a particular measurement tool is developed. In general, one can ascertain that reliability, validity and responsiveness are the most critical measurement properties (Mokkin, et al., 2010; Carter, et al., 2011; de Vet, et al., 2011) (figure 3.14). Any instrument to be used for the purpose of measuring a health-related issue should be subject to validity and reliability testing. To be able to deploy this device for clinical use, we first need to test its reliability, validity and responsiveness to a certain degree (Mokkink, et al, 2010).

Figure 3.14 – Measurement properties of outcome measurement instruments
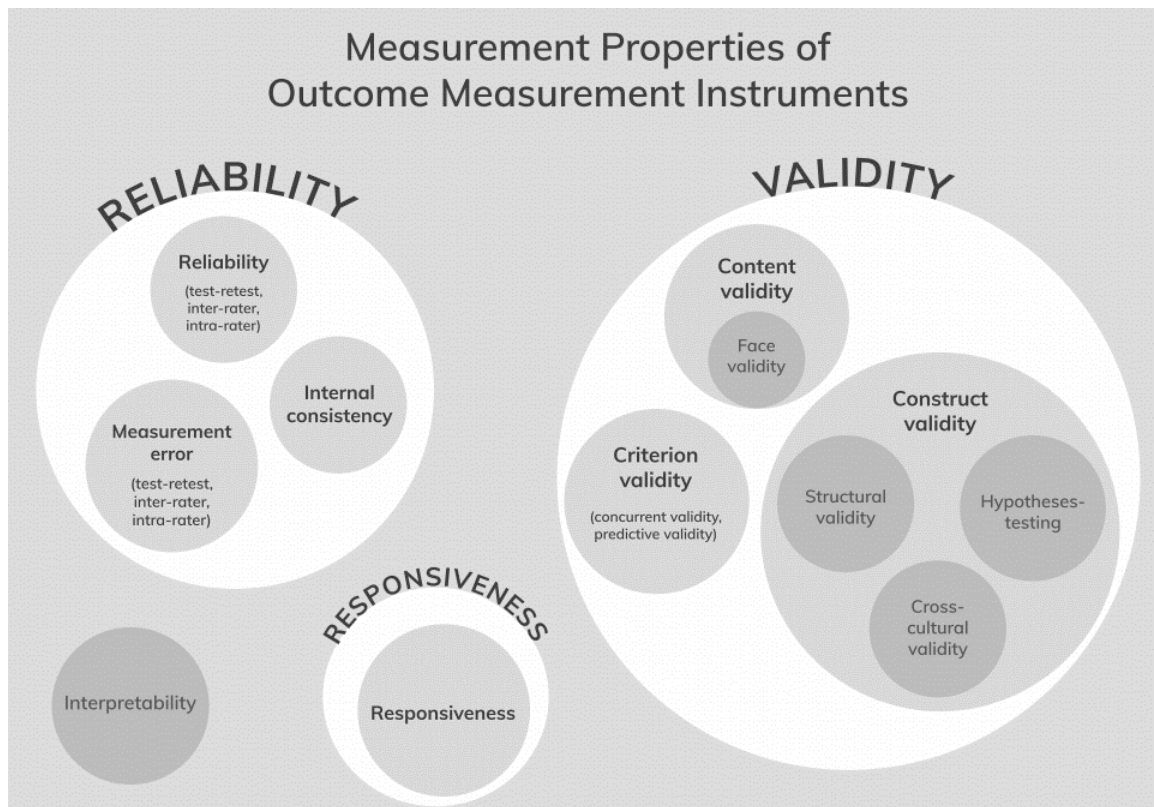
Source: *Mokkink, et al. (2010)*

3.5.1   Reliability

Lexel and Downham (2005) explained that "reliability refers to the reproducibility of measurements" while Mokkink, et al (2010) defined reliability as "the degree to which the measurement is free from measurement error". In a more extended way, the same authors also add that "the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same multi-item measurement instrument (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)". Due to the instrument's characteristics in this work, data collection focused on intra-rater and inter-rater reliability. As suggested by Sharma, et al. (2014) test-retest for Intrarater reliability were done on different days rather than on the same day, and the mean of three repetitions was used for reliability comparison. Between day data used the average of three

repetitions for comparison – ICC(2,3). If between day data was not available, comparison between each of the three repetitions was performed - ICC(2,1).

3.5.2   Relative reliability

Reliability can then be split into relative reliability and absolute reliability. Relative reliability implies that the same score measured on two different occasions (repeated measure) will have a similar overall ranking as it relates similarly to the other scores in both measurements. This type of reliability is assessed by correlation coefficients, which measure the relationship when comparing two or more output measures. The commonly used measure for correlation coefficient in HHDs is the intraclass correlation coefficients (ICC) which will be applied to all data collection procedures. The ICC varies between 0 and 1, with 0 meaning no correlation and 1 meaning a perfect correlation (Carter, et al., 2011). Fleiss (1999) described reliability, with an ICC over 0.75 considered excellent, between 0.4-0.75 as fair to good and poor if below 0.4. More recently, Portney and Watkins (2015) used an updated version with a stricter view in regards to poor reliability, defining poor as anything below 0.5, between 0.5 – 0.74 as moderate reliability, above 0.75 as good reliability, and an ICC larger than 0.90 as excellent – this version was selected to report the data.

3.5.3   Absolute reliability

Absolute reliability refers to the variability of the same score over a repeated measurement - in other words, it is the expected instrument error over two different measurements. This is assessed by the SEM, which measures the standard deviation for a specific group of measurements (Carter, et al., 2011). The author used SEM, which uses one standard deviation as a multiplier which regards to an SEM that covers 68% of the variability, this is the most common method of reliability, although some authors have used SEM covering two standard deviations or 95% SEM.

3.5.4   Validity

Validity can have several categories (Carter, et al., 2011, Portney and Watkins, 2015): 1) Construct validity refers to the abstract construct concerning the measurement being taken; this type of construct compares different populations to assess if the new measure can differentiate between them. 2) Content validity attempts to determine if a tool can represent the concept under study; this is usually used for self-reported or observational tools. 3) Criterion validity "is the extent to which one measure is systematically related to other measure or outcome". It is commonly divided into concurrent validity, which compares the new tool with the gold-standard measure (investigated in this thesis), and predictive validity, which looks into a test taken at a point in time trying to foresee a future result.

To assess how the measures or outcomes relate, correlation tests are usually performed (Carter, et al., 2011, Portney and Watkins, 2015). For instance, the correlation coefficient (or Pearsons's *r*) has been used extensively to measure association. However, the approach has its problems with the main caveat being pointed by several authors because a strong association between measures might not reflect essential differences between methods (Bland and Altman, 1983, Daly and Bourke, 2000). This is why a different approach was proposed by Bland and Altman in 1986, which is known as the Bland-Altman Analysis (BA Analysis). Their work is one of the most commonly cited papers when comparing methods in agreement research to validate one method against another in several domains (Scolletta, et al., 2016; Lynall, et al., 2017), as it allows for further insight onto differences between methods across the spectrum of measurements. An important concept of BA analysis is the pre-setting of the Limits of agreement (LOA), this helps researchers set an acceptable level of differences between the two methods being tested. In this work, LOA were not set *a priori* as it is usually recommended (Giavarina, 2015; Abu-Arafeh, Jordan and Drummond, 2016) for three main reasons. Firstly, when using the BA analysis, researchers in several scientific domains are looking for the possibility of replacing one measure with another, however, in HHD research, authors are trying to quantify the differences between the gold standard and a cheap/widespread option which is the HHD. HHDs, where testers deploy resistance (such as ASSA), will most likely never replace IKDs due to the introduction of human error (technique, strength, motivation, etc.) and therefore, the exercise of setting a priori LOA

to establish if the new method is adequate to replace the gold-standard seems redundant. Secondly, using a concentric approach with an HHD has only been done once (Le-Ngoc and Janssen, 2012), and there is insufficient data from previous research to set LOA appropriately. Lastly, similar HHD research investigating make or break tests for validity does not always use BA analysis (some authors only use correlation coefficients). If they use it, data is not always reported in the same units, which means there is also insufficient consensus to establish useful LOA.

More recently, another approach has been considered to complement the BA analysis which was proposed by Ludbrook (2010). In that paper, the author suggested that a Model II regression should be performed. In a subsequent paper Ludbrook (2012) suggested the use of Weighted Least Products (WLP) – as it was more appropriate for method comparison studies, particularly if one is attempting to calibrate a device against the gold standard.

Therefore, and because there are no clear guidelines in the literature for one specific approach: the correlation was measured using Pearson's r (recommended by Ludbrook (2010) as part of the Model II regression analysis) with classed as small (r<0.30), moderate (0.30-0.50), large (0.50-0.70), very large (0.70-1.0) (Hopkins, et al., 2009); whereas agreement was measured using Bland-Altman analysis – qualitative assessment (Bland and Altman, 1986); to provide a quantitative assessment and a calibration equation Weighed Least Products (WLP) regression was also used to assess validity (Ludbrook, 2010). With the WLP regression, the assessment of validity is provided by the 95% confidence intervals (CI), from the regression equation, if the 95% CI for the intercept does not include 0, then fixed bias can be assumed, in the same way, if the 95% CI for the slope does not include 1, proportional bias can be assumed (Ludbrook, 2002).

3.5.5   Responsiveness

When developing a new tool, practitioners are also interested in knowing how it can detect changes to treatment or the pathology's normal progression as this will impact the rehabilitation process. This means end-users are interested in identifying the real changes to their patient's status that are not measurement errors. Responsiveness is given by Minimal Detectable Difference (MDD) or Minimal Detectable Change (MDC) and is

related to SEM due to its statistical formula MDD = z * SEM * √2 and with the value of z set for 95% confidence interval equal to 1.96 (Portney and Watkins, 2015). It can be defined as the least amount of change that the measurement tool can detect while exceeding measurement error and is an important value to determine the clinical relevance and usability of a measurement (Carter, et al., 2011). SEM and MDD values were only retrieved if ICC values were higher than 0.50.

These are the variables that will provide the baseline information regarding the performance of ASSA while using active resistance manual muscle tests. Further information regarding statistical analysis and category-specific formulas can be found in the next chapters.

3.6 Aims

The aims of this thesis are:

a) To create and develop a new hand-held dynamometer for concentric muscle tests.
b) To investigate the validity, reliability and responsiveness of the new prototype and subsequent versions in a healthy population, by using groups of testers with both homogeneous and heterogenous experience levels.
c) To develop and investigate the use of sonification as an adjunct of manual muscle test when using a hand-held dynamometer.
d) To identify how the use of a new approach for manual muscle tests using a newly developed prototype is perceived by physiotherapists and how it can further improve for dissemination and implementation in clinical practice.

Hypothesis for validity, reliability and responsiveness testing are presented in Appendix 6 for each chapter.

4. A PILOT STUDY EVALUATING CONCURRENT VALIDITY AND RELIABILITY OF A NEW HAND-HELD DYNAMOMETER ON HEALTHY SUBJECTS

4.1 Introduction

The previous chapters document the use of sonification in physiotherapy and its limitations in assessment and intervention approaches. Simultaneously, the author demonstrated how current strength assessment practices are still far from ideal in several settings where IKDs or HHDs are not available. This highlights the need for alternative methods of strength testing to improve the current paradigm where physiotherapists still rely on MMTs. This chapter presents the pilot testing of an Arduino-based Sound Strength Assessment (ASSA) tool by investigating concurrent validity and reliability in healthy subjects.

4.2 Hand-Held dynamometers and ASSA

As reported in chapter 3, ASSA v2 is a new portable HHD prototype designed for strength testing to provide a cheaper, quantitative and concentric approach to physiotherapists and exercise professionals. The current study investigates its ability to provide comparable data to a gold-standard technique (validity) and between users (reliability) without sound-feedback.

Even though the ASSA v2 can be considered in the same category as HHDs, it has the potential to provide extra information such as range of motion (ROM), peak force and angle of peak force, variables that are not commonly found in a standard HHD. Typically, an HHD is a device which uses a load cell and a small processor to record muscle strength data from an isometric activity in several different joints. Examples of devices available on the market range from the MicroFET™ 3, Lafayette™ and PowerTrack™ II Commander.

Validity and reliability problems are key measures for assessing a new device. In this chapter ASSA v2 function is compared with a gold standard method (IKD) for validity and reliability (intra and inter-tester) purposes.

4.3 Objective

The pilot study's objective was to investigate ASSA v2 validity and reliability in a laboratorial scenario by comparing ASSA v2 with a gold standard measure (IKD). The aim was to examine the device's performance and explore any problems related to its use in a real-life scenario.

4.4 Methods and materials

4.4.1 Design

A repeated-measures, single-blinded randomised controlled trial was performed with a healthy adult population. ASSA v2 was tested. Three different testers with different degrees of experience and in two different sessions. Participants were randomly allocated to each intervention and tester (Day 1 - IKD and ASSA; Day 2 – ASSA Tester 1 and ASSA Tester 2/3) (figure 4.1) to minimise bias, and the order of movements executed by each tester/device was maintained. Tester 1 was an experienced physiotherapist with experience in using HHD and manual muscle tests. Tester 2 and tester 3 had no experience with manual muscle tests or HHD.

Figure 4. 1Research outline by day

4.4.2   Recruitment and participants

Participants consisted of staff and students from ARU. The ARU website, posters and word-of-mouth were used for recruitment. The SES School Research Ethics Panel approved the research from ARU – Cambridge with the code ESPGR-02 (Appendix 7).

The sample size was determined by considering a power of 90% and alpha of 0.05 according to the work of Bujang and Baharum (2017), with a difference of acceptable reliability (0.70) to expectable reliability (0.90), this study would need to recruit at least 18 individuals with three repetitions per participant. Considering some data may be lost, and participants may withdraw from the study; it was thought more appropriate to recruit 20%-30% more participants. In the end, a total of 24 subjects were recruited for the study.

Individuals who demonstrated interest in participating in the study were provided with a Participant Information Sheet (PIS) and a Participant Consent Form (PCF) which was signed before taking part in the study. The procedure was explained, and a questionnaire was completed to apply exclusion and inclusion criteria.

Inclusion criteria were: between 18 – 40 years old; no major surgery/injury in the last 6 months; no current injury on back or lower limbs. While exclusion criteria were set

as: Currently pregnant; Hearing difficulties; Recent injury or disability that prevents the subject from performing a maximum voluntary contraction.

Testers for the use of the device were also recruited from ARU. The data collection was initially planned with two testers (one experienced – tester 1 and one novice – tester 2) However, after data collection had started, the novice tester had to withdraw due to personal issues. A third tester (tester 3) was recruited to evaluate the remaining participants.

4.4.3   Testing procedure

This protocol aims to control factors that could influence testing, such as stabilisation, testing position and verbal feedback by providing a procedure that can be used transversely to diminish bias and optimise reliability.

Before starting the protocol (either when using IKD or ASSA) each participant initiated a 5-min warm-up on a cycle ergometer using a braking force corresponding to 2% of their body mass at 60 revolutions per minute. The familiarisation procedure was performed just before the correspondent movement was done in each device. These steps aimed to facilitate the testing procedure by providing the first contact with the instruments. The participant was allowed to do one to three 50% - 60% Maximal Voluntary Contraction (MVC) until they felt comfortable with the device before initiating the test.

The familiarisation trials facilitated the subsequent tasks as they allowed participants to perform the movement adequately and gave the testers the chance to adjust the force applied in each movement. Force should be exerted on ASSA at a progressive and slow rate (1-2 seconds are required for the tester to adequate the resistance level and device positioning), regardless of the testing being isometric or not. In the isometric testing, the participant was asked to maintain the contraction for 5 seconds (this feedback was provided not by the tester using ASSA but by the other tester present in the room). A goniometer was used to determine the joint angles at which each movement is deemed to occur (isometric movements), namely at 60° knee extension or 90° knee flexion.

This familiarisation might be perceived as reduced when compared with what other authors used, such as Whiteley, et al. (2012) - three sub-maximal contractions followed by three near-maximal concentric contractions at the same speed and mode as the test about to be conducted. After pre-trial testing, it was concluded that such an approach,

Whiteley, et al. (2012), would be deemed too fatiguing as the participants are not professional athletes and this study is considerably longer in terms of protocol. As required in studies of this nature, the encouragement was standard and firm to allow for the best possible results as it has been demonstrated in the literature that this can affect the outcome (Engel, et al., 2019; Rendos, et al., 2019). The same information was used in ASSA and IKD testing by saying to the participant: "When I say three start pushing against the device, start slowly to remove the slack between the device's foam and your leg and then push as hard as you can". The encouragement was consistent on both devices and consisted of "Push, push, push and stop" or "Bend your knee, bend, bend and stop". These instructions were kept over chapter 5 and 6.

4.4.4   Manual muscle test protocol

Participants were required to perform knee extension (Figure 4.2) and knee flexion (Figure 4.3) movements in two different positions on two different days in consecutive weeks.

Figure 4. 2 Knee extension
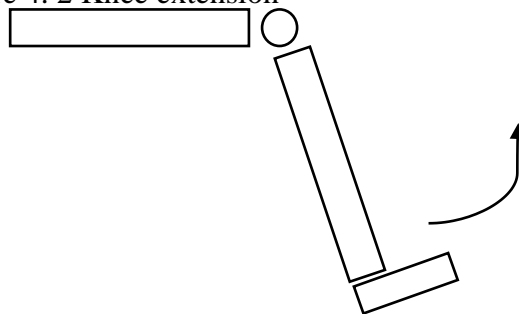
Figure 4. 3 Knee flexion
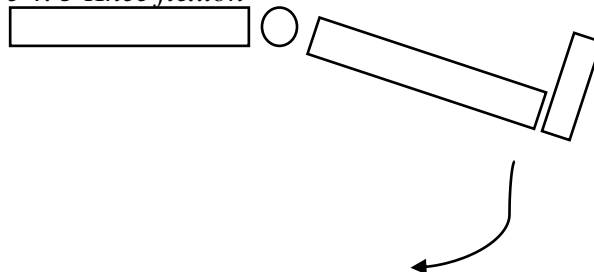
To address this, inexperienced testers had a previous two-hour training session on the theoretical background of the use of manual muscle testing and the use of ASSA with an experienced physiotherapist. Testers were provided with practical training on the use of manual muscle tests. The author was responsible for both concentric and isometric training by executing the tests with and without the HHD and providing adequate verbal feedback throughout the procedure. Testing set-up is demonstrated by figure 4.4.

Figure 4. 4 ASSA prototype and set-up

Testing was done with at least a five-minute rest between trials to reduce fatigue, similar to the protocol developed by Mentiplay, et al. (2015). Between each testing day there was a 5-12 day interval, as per other studies (Kilmer, et al., 1997; Thorborg, et al., 2010; Stockton, et al, 2011; Dowman, et al., 2016), to minimise any learning effect but still guarantee that there are no considerable changes in strength due to training.

Participants were given 5-10 seconds to rest between contractions, 1 minute between movements and 5-10 minutes between devices/testers and for repositioning to reduce fatigue (Whinton, et al., 2018). Participants were also asked not to perform any vigorous physical activity in the 48h preceding each testing session and to maintain their nutritional and activity habits. Subjects were, as much as possible, scheduled to perform the test at similar times of the day to minimise any diurnal influence in strength production. The participants used the same warm-up routine on both the IKD and manual muscle testing days, and all testing was performed on the dominant leg (this was determined beforehand by asking the participant which leg they used to kick a ball).

In HHD research, make tests are usually preferred as they are easier to perform due to the lower forces at play when compared to break tests (Bohannon, 1988; Stratford and Balsor, 1994). However, since they are isometric, there is no information about how force changes through the range of motion.

As the new ASSA prototype can collect information through the range of motion, it was decided to use isometric testing and the technique described by the Oxford Scale, also named Medical Research Council Manual Muscle Testing scale. The assessment is made from resisting movement through the available range of motion using a concentric movement (concentric testing) in this type of testing. For the isometric (figure 4.5 and 4.6) and concentric (figure 4.7) testing, the test's position was standardised (Table 4.1) for both tester and participant.

As per the Oxford Scale, manual testing requires maximum concentric force to be resisted in the available ROM. While in the isometric testing, participants were asked to exert as much force as possible for 5 seconds to elicit maximal muscle fibre recruitment (Bohannon, 1997). The two testers were always present for testing (one performing the MMTs; the other, providing support and controlling the laptop). In order to guarantee a specific position for the isometric test, a goniometer was placed by the assistant to indicate the correct initial position. Any loss of control throughout the range or inability to keep the isometric movement's predetermined position meant the test was repeated. If the tester was still unable to conduct a valid test, then this would be categorised as "invalid" and the data discarded.

Figure 4. 5 Isometric knee extension

Figure 4. 6 Isometric knee flexion

Regarding the knee flexion movements, it was decided to standardise this movement by asking participants to "keep their foot and toes up" during ASSA and IKD testing (this was also performed on chapter 5 and 6). This is due to the influence of the gastrocnemius on knee flexion moment torque which is minimised in this position (Galluci and Challis, 2002).

Figure 4. 7 Knee flexion concentric testing position

Table 4. 1 Testing position by movement

Activity	Subjects' Position	ASSA
Knee Extension Isometric and concentric	Sat on an IKD chair, cross the arms and not hold the testing chair during the procedure. The chair was upright at 85°, and back support was adjusted to allow free knee flexion. **Isometric extension** - the test starts at 60° knee extension **Concentric extension** – the test starts from a relaxed seated position (+-90/100°).	The tester directed ASSA's contact area to the anterior lower third of the tibia just above the tibial malleoli while maintaining a gricular orientation with the lower leg segment, maintained throughout the movement.
Knee Flexion Isometric	The subjects were asked to sit on an IKD chair, cross the arms and not hold the testing chair during the testing procedure. The chair was upright at 85°, and back support was adjusted to allow free knee flexion. **Isometric extension** - the test starts at 90°.	The tester directed ASSA's contact area to the posterior lower third of the tibia just above the tibial malleoli while maintaining a perpendicular orientation with the lower leg segment, maintained throughout the movement.

Knee Flexion concentric	Prone position, on a plinth, while keeping their arms next to their body and not hold the testing plinth during the testing procedure, lower limbs straight and toes hanging over the edge of the table. Allow for at least 5-10° of knee flexion before applying resistance to facilitate the application of force to the segment.	The tester directed ASSA's contact area to the posterior lower third of the tibia just above the tibial malleoli while maintaining a perpendicular orientation with the lower leg segment, maintained throughout the movement. Position of Therapist: On the side of the leg to be tested. The dynamometer on the posterior surface of the leg just above the ankle.

The testers were instructed to use a modified technique from Manual Muscle Testing (Hislop and Montgomery, 2007) with an appropriate body position to minimise testing errors and optimise mechanical advantage over the participants (Kelln, et al., 2008).

Throughout testing the testers directed ASSA's contact area to the lower third of the tibia (anteriorly or posteriorly according to the movement tested) and a 90° orientation with the moving segment was enforced throughout the movement to maximise force delivered and facilitate resistance. Testing positions used followed the protocol proposed by previous authors such as Bohannon (1986); Dowman, et al. (2016) and Muff, et al. (2016). At the end of the protocol, a 5-10 minutes rest was provided between testers or before changing to the IKD. The protocol for HHD is described in table 4.2.

Table 4. 2 Protocol outline - ASSA

Muscle group	Activity (3x)	Starting position	Time	Position	Rest
Knee extensors	Isometric	60°	5 seconds	Seated	1 min.
	Concentric Extension	Relaxed position	N/A	Seated	1 min.
Knee Flexors	Isometric	90°	5 seconds	Seated	1 min.
	Concentric Flexion	Relaxed position	N/A	Prone	5-10 min.

Note: N/A – Not applicable; min.: Minute

4.4.5   IKD Protocol

The IKD (HUMAC NORM Model 770; CSMi, Stoughton, MA, USA) protocol followed the manufacturer's instruction for this type of activity. All movements were recorded using the gravity corrected feature. The chair was set-up in two different positions, seated and prone. For the seated position, the tester adjusted the chest, thigh, and lower leg straps. The subject was then asked to cross the arms and not hold the testing chair

during the procedure. The subject was positioned as per the manufacturer's instructions with a chair back angle of 85°, and the other parameters were adjusted to match individual differences. In the prone position the back of the seat was positioned at 0° so the subject would be flat with the alignment of the axis of the dynamometer in the same region as for the seated position. For further details see table 4.3.

The knee joint axis was aligned with the mechanical axis of the dynamometer. About 2cm were allowed between the back of the tester's leg and the chair to allow for maximal flexion. Each participant's chair position was recorded so it could be positioned in the same orientation for the follow-up trial (The IKD chair was used for all seating tests). The participants' arms were kept on the side without holding the chair. For the prone position, a plinth was used, and no straps were utilised.

Table 4. 3 Protocol outline - IKD

MUSCLE GROUP	ACTIVITY (5X)	STARTING POSITION	SPEED	POSITION	REST
KNEE EXTENSORS	Isokinetic	Knee flexion (90º-110º)	30°/sec	Seated	1 min.
	Isokinetic	Knee flexion (90º-110º)	60°/sec	Seated	1 min.
	Isometric	60º flexion	N/A	Seated	1 min
KNEE FLEXORS	Isometric	90° flexion	N/A	Seated	1 min.
	Isokinetic	Knee flexion (110º-120º)	30°/sec	Prone	1 min.
	Isokinetic	Knee flexion (110º-120º)	60°/sec	Prone	5-10 min.

On average, 60 seconds rest was provided between sets to allow for recovery and in line with previous research (Parcell, et al., 2002). Isometric knee extension was assessed at 60° degrees (0° – full extension) as this was reported to be the mean angle of peak force (Deones, Wiley and Worrell, 1994) while isometric knee flexion was tested at 90° to be similar to the test done by Andrews, Thomas and Bohannon (1996). The IKD tested each participant maximal available range for each movement.

4.5   Data processing

The author wrote a customised RStudio (Version 1.2.5033 © 2009-2019) program to analyse the data collected from the ASSA and the IKD.

Data processing consisted of the following steps:

1. A zero-phase shift 5Hz low-pass 4th order Butterworth filter was designed to filter the data from the devices. As the devices had different sampling rates (ASSA – 10Hz; IKD – 500Hz), a cubic spline interpolation was used to allow for a more accurate comparison.

2. ASSA measures strength in Kg, unlike the IKD, which uses Newton-meter (Nm) as the standard unit of measure. Taking this difference into account, and for validity purposes, it was necessary to convert ASSA strength from Kg to Nm. To do this, Kg were transformed into Newtons (1Kg = 9.81N), but lower leg segment sizes were also needed. As those were not gathered at the time of data collection, the height estimation formula from Trotter and Gleser (1958) was used to estimate individual leg segment size using a linear regression analysis. The final value of the leg segment used for torque estimation was equal to the theoretical values obtained from Trotter and Gleser (1958) minus 4cm to account for the dynamometer's position on the participant's lower leg. Data in Nm was then compared.

3. Gravitational moment correction for the ASSA force output (when comparing the IKD) was done using the method suggested by Kellis and Baltzopoulus (1996), the authors used anthropometric data from Dempster (1955). Following their equation the knee joint gravitational correction was defined as: M = (l*0.437)*(0.06*BW), with M =

gravitational moment, l = length of the limb (m), 0.437= centre of mass proximal joint distance/segment length, 0.06 = leg-foot weight/body weight, and BW = body weight (N).

4. Data were tested for normality using Shapiro-Wilks (p<0.05). Mean and standard deviations (SDs) are reported. Muscle strength was reported in kilograms (Kg) for ASSA when comparing values between testers but was reported in Newton-meter (Nm) when being compared with the IKD output (Nm). The angle of peak torque results are reported in degrees (°). For both types of reliability results were categorised as poor if below 0.5, moderate if between 0.5 – 0.74, as good if above 0.74, and as excellent if larger than 0.90.

5. A dependent t-test was performed to compare the mean force for each movement tested, effect sizes are reported with Cohen's D with small effect reported as r = 0.10, medium effect as r = 0.30 and large effects as r = 0.50. Bonferroni correction was used when performing the analysis with a p-value of 0.025.

4.5.1   Data analysis

Intra-rater reliability and responsiveness: The approaches to analyse Intrarater reliability included intraclass correlation coefficient (ICC), standard error of measurement (SEM) and in percentage (SEM%) while Minimal detectable change (MDC) in percentage (MDC%) was used to test for responsiveness. The degree of correlation as defined by Schrout and Fleiss (1979) is calculated using a mean of three repetitions and comparing the data between different days of all the maximal strength tests and is calculated using ICC (2,3). The ICC (2,3), uses a two-way random effects model. Results also include reliability measures (intra-day) for comparison between three repetitions for each tester– ICC (2,1) (Koo and Li, 2016).

Inter-rater reliability and responsiveness: To assess the degree of reliability between different testers, the tests used were similar to Intrarater reliability: ICC, SEM and MDC. The degree of correlation between testers was done using the average (force or angle) of three repetitions of all the maximal strength tests. The ICC (2,3) reflects a two-way random effects model.

SEM results were obtained by multiplying the standard deviation from the first session output (SD_1) by the square root of one minus the ICC:

$$SEM = SD_I\sqrt{1 - ICC}$$

This was then converted to a percentage (SEM%) to facilitate comparison with similar research following the equation:

$$SEM\% = (SEM/mean) \times 100$$

(Portney and Watkins, 2015)

MDC for 95% confidence intervals were calculated with $z = 1.96$:

$$MDC = z \times SEM \times \sqrt{2}$$

This was also reported in percentage (MDC%):

$$MDC\% = (MDC/mean) \times 100$$

(Lexell and Downham, 2005)

These tests were done for the maximum force produced (peak force) and angle of which this peak force was achieved (angle of peak force), to facilitate comparison with similar research on both inter and intratester reliability (Portney and Watkins, 2015).

Concurrent Validity: To assess validity of the new prototype; Pearson's r correlation, Bland Altman analysis – Limits of Agreement (LOA). (Zaki, et al, 2012). WLP regression was used to quantify the agreement between the two devices and can provide a calibration equation (Ludbrook, 2010). Data analysis followed these steps:

1. To determine concurrent validity Bland Altman analysis were run for each movement assessed using the prototype (only for tester 1 who assessed all participants). Comparisons with the IKD used knee extension and flexion data at 30º/sec, 60º/sec and isometric.

2. Shapiro-Wilk (p <0.05) test was used to assess normality. When the data set was determined to be non-normal (line of best fit different from zero) a log-transformed BA analysis was done.

3. Due to difficulty to interpret Bland-Altman plots on a log scale, results were transformed back to the original scale using anti-logs and plotting the data from the ratio between ASSA and the IKD against means from both devices (Bland and Altman, 1999).

4. If the line of best fit from the BA analysis was still not zero after the log transform, then a regression-based V-shaped BA plot was done (Ludbrook, 2010). To aid interpretation of this plot a percentage BA plot was done as in the above point. BA plots were not created to determine a numeric output that determines if a method is valid or not compared to another but to give an indication of differences between methods given a determined spectrum of measurements.

4.6   Results

A total of 24 participants were recruited (6 females and 18 males) between 18 and 40 years of age mean (SD) - 24.92 (6.18). Mean (SD) height was 174.82 cm (8.43) and ranged from 158cm to 188cm; mean (SD) weight was 77.58kg (20.27) with a range of 55kg to 150kg. Participants were required to attend two sessions, first for IKD and ASSA testing and on the second day for ASSA testing with two different testers. Two participants did not return for the second assessment, and therefore no data was collected regarding inter-tester reliability. In total, tester 1 assessed 24 participants at least once, tester 2 assessed 12 participants, and tester 3 assessed 10 participants.

Tester 1 could not resist isometric knee extension movement from one of the participants whereas Tester 2 could not resist knee extension isometric movement in 6 participants or resist/maintain control during knee extension concentric movement for 3 participants. Lastly, tester 3 could not maintain position with 3 different participants when assessing

isometric knee extension. For each movement that testers could not resist, data was discarded and not used for statistical analysis.

4.6.1 Intra-tester reliability

4.6.1.1 Peak Force

For between day reliability (day 1 vs day 2), the ICC (2,3) results (table 4.4) demonstrated increased reliability for peak force using the knee extension concentric test compared to the extension isometric data and equal reliability when comparing concentric flexion with isometric flexion. For tester 1, ICC was 0.93 (95% CI: 0.8-0.97) for knee extension concentric when compared with isometric extension 0.83 (95% CI: 0.6-0.94). Whereas knee flexion concentric ICC was 0.83 (95% CI: 0.6-0.93) and isometric knee flexion was 0.83 (95% CI: 0.6-0.93).

Table 4. 4 ICC(2,3), SEM and MDC for peak force per movement tested (day 1 vs day 2 – Tester 1)

	ICC (CI)	SEM (%)	MDC (%)
ISOM EX	0.83 (0.6-0.94)	3.48 kg (5.9%)	9.65 kg (20.4%)
CONC EXT	0.93 (0.8-0.97)	2.12 kg (5.04%)	5.88 kg (14%)
ISO FLEX	0.83 (0.6-0.93)	2.72 kg (9.8%)	7.54 kg (27.2%)
CONC FLEX	0.83 (0.6-0.93)	1.80 kg (8.9%)	4.99 kg (24.74%)

Note: ISOM EX- Isometric extension; CONC EXT- Concentric extension; ISO FLEX- Isometric flexion; CONC FLEX- Concentric flexion; CI – Confidence Interval

The concentric tests display smaller SEM (1.80 and 2.12kg for flexion and extension respectively) when compared with isometric testing (2.72 and 3.48kg for flexion and extension respectively) (see table 4.4). MDC results show that concentric testing (4.99kg

for flexion and 5.88kg extension) is smaller than isometric testing (7.54kg for flexion and 9.65kg for extension).

Total mean force production for tester 1 comparing isometric and concentric tests (table 4.5) reveals that isometric tests produce on average more force than concentric tests regardless of the day and tested articulation. Table 4.6 shows statistically significant higher force produced by isometric testing on extension (Day 1 – 4.94kg and Day 2 – 7.17kg ($p < 0.001$)) and knee flexion (Day 1 – 7.22kg and Day 2 – 8.15kg ($p < 0.001$)) on average (table 4.6), large effects sizes for the difference between isometric and concentric tests were also found.

Table 4. 5 Strength mean in Kg(SD) for day 1 and day 2 (Tester 1)

	ISOM EX	CONC EXT	ISO FLEX	CONC FLEX
DAY 1	47.61 (8.46)	42.67 (8.44)	26.53 (6.61)	21.20 (4.40)
DAY 2	49.84 (10.01)	42.67 (9.84)	27.36 (8.04)	21.55 (5.20)

Note: ISOM EX- Isometric extension; CONC EXT- Concentric extension; ISO FLEX- Isometric flexion; CONC FLEX- Concentric flexion

Table 4. 6 T-test comparison of mean force differences isometric vs concentric tests by day and joint

	DIFF	95% CI	T - VALUE	P-VALUE	EFFECT SIZE
ISOM. VS CONC. EXTENSION – D1	4.94	3.02 – 6.68	5.33	<0.001	0.75
ISOM. VS CONC. FLEXION- D1	7.22	5.26-9.19	7.61	<0.001	0.85
ISOM. VS CONC. EXTENSION – D2	7.17	4.75-9.59	6.18	<0.001	0.81
ISOM. VS CONC. FLEXION- D2	8.15	4.30-12	4.40	<0.001	0.69

Note: ISOM- Isometric; CONC- Concentric; D1 – Day 1; D2 – Day 2

For the intra-session reliability (ICC(2,1)) the following tables 4.7 and 4.8 describe results for each movement (either the maximum force or angle of peak force output). Table 4.7 displays good reliability (ICC≥0.80) for tester 1 in both extension and flexion concentric tests (peak force) with an SEM of 6.78% and 9.51%, respectively. Tester 2 showed good reliability for both extension and flexion (table 4.8), with an ICC of 0.85 and 0.87 respectively and SEM of 5.13% and 9.49%. For tester 3, the reliability was poor with an ICC of 0.36 (extension) and 0.44 (flexion).

Table 4. 7 Intra-session reliability ICC (2,1), SEM and MDC for tester 1 (Day 1)

T1	ICC (2,1)	P-VALUE	CI	SEM (%)	MDC (%)
CE	0.89	<0.01	0.80-0.95	2.87 (6.78%)	7.96 (18.8%)
CF	0.80	<0.01	0.65-0.90	1.90 (9.51%)	5.3 (26.41%)

Note: CE - Concentric Extension; CF - Concentric Flexion

Table 4. 8 Intra-session reliability ICC (2,1), SEM and MDC for tester 2 and 3

T2	ICC (2,1)	P-VALUE	CI	SEM	MDC
CE	0.85	<0.01	0.67-0.95	1.49(5.13%)	4.13(14.25%)
CF	0.87	<0.01	0.69-0.95	1.63(9.49%)	4.52(26.34%)
T3	**ICC (2,3)**	**p-value**	**CI**	**SEM**	**MDC**
CE	0.36	<0.05	0.00071-0.75	NA	NA
CF	0.44	<0.01	0.064-0.79	NA	NA

Note: CE - Concentric Extension; CF - Concentric Flexion

4.6.1.2   ***Angle of Peak Force***

Angle of peak force for tester 1 (table 4.9) shows moderate reliability (between days 1 and 2) for knee extension (ICC – 0.61). Knee flexion revealed moderate reliability as well with ICC=0.71. The SEM(%) for knee extension was 9.28%, and an MDC(%) of 25.70%; a larger SEM and MDC for knee flexion was found with 18.52% and 51.29%, respectively.

Table 4. 9 Between session comparison for tester 1 angle of peak force

T1	ICC (2,3)	P-VAL.	CI	SEM(%)	MDC(%)
EXT.	0.61	<0.05	-0.001 – 0.85	3.01(9.28%)	8.34(25.70%)
FLEX.	0.71	<0.01	0.34– 0.86	5.82(18.52%)	16.13(51.29%)

Note: Ext. – Extension: Flex. - Flexion

When comparing ICC(2,1) levels for angle peak force the results in table 4.10 reveal poor reliability for tester 1 on both day 1 (0.23) and 2 (0.48) when resisting knee extension, while tester 2 (0.53) and tester 3 (0.55) show moderate reliability. Knee flexion reliability levels increase slightly for both tester 1 – day1 (0.59), tester 1 - day 2 (0.68) and tester 3 (0.59), while tester 2 (0.43) results were poor. SEM(%) values from knee extension concentric was 15% (tester 2), meanwhile knee flexion ranged from 21.63% (tester 1 – day 1), 27.45% (tester 1 – day 2), and 35.38% for tester 3. MDC(%) values for tester 1 were quite high for knee flexion at almost 60% for day 1 and over 70% at day 2. Novice testers 2 and 3 demonstrated similar values with 41.72% and 52.68% respectively for knee extension. MDC(%) for tester 1 displayed an MDC higher than 59% in both days and tester 3 with an MDC of 97.99%.

The results from the reliability for the IKD (ICC(2,1)) are reported in table 4.11 with moderate to good reliability in all movements tested at 30º/sec and 60º/sec. MDC values were lower at 60º/sec than at 30º/sec.

Table 4. 10 Intra-tester ICC(2,1) Angle of peak force (º) - ASSA

T1	ICC (2,1)	P-VALUE	CI	SEM(%)	MDC(%)
EXTENSION – D1	0.23	0.03	-0.01 – 0.51	NA	NA
FLEXION – D1	0.59	<0.01	0.37 – 0.78	6.53 (21.63%)	18.10 (59.93%)
EXTENSION – D2	0.48	<0.01	0.23 – 0.71	NA	NA
FLEXION – D2	0.68	<0.01	0.48 – 0.84	6.64 (27.45%)	18.40 (76.05%)
T2	**ICC (2,1)**	**p-value**	**CI**	**SEM(%)**	**MDC(%)**
EXTENSION	0.53	<0.01	0.17 – 0.82	6.26 (15%)	17.34 (41.72%)
FLEXION	0.43	<0.05	-0.032 – 0.85	NA	NA
T3	**ICC (2,1)**	**p-value**	**CI**	**SEM(%)**	**MDC(%)**
EXTENSION	0.55	<0.01	0.17 – 0.85	6.49 (19.01%)	17.97 (52.68%)
FLEXION	0.59	<0.01	0.21 – 0.86	8.73 (35.38%)	24.18 (97.99%)

Table 4. 11 Intra-tester ICC (2,1)Angle of peak force (º) – IKD

IKD 30°/SEC	ICC (2,1)	P-VALUE	CI	SEM(%)	MDC(%)
EXTENSION	0.76	<0.01	0.53-0.89	5.16 (8.19%)	14.30 (22.70%)
FLEXION	0.85	<0.01	0.7-0.93	8.08 (18.79%)	22.39 (52.07%)
IKD 60°/SEC	**ICC (2,1)**	**p-value**	**CI**	**SEM(%)**	**MDC(%)**
EXTENSION	0.84	<0.01	0.68-0.93	3.19 (7.38%)	8.84 (20.43%)
FLEXION	0.70	<0.01	0.42-0.87	3.53 (5.72%)	9.77 (15.85%)

4.6.2   Inter-tester reliability

4.6.2.1   **Peak Force**

As seen in table 4.12, ICC(2,3) tests display a good correlation for inter-rater reliability in the Isometric Flexion T1/T2 movement. Reliability is considered moderate in Isometric Flexion (T1/T3) and concentric Flexion (T1/T3). All remaining comparisons show poor inter-tester reliability.

Table 4. 12 ICC (95% CI) peak force between testers

T1/T2	ICC (2,3)	P-VALUE	CI	SEM(%)	MDC(%)
CE.	0.18	0.3	-0.3 – 0.3	NA	NA
ISO. EXT.	0.32	0.24	-0.24 – 0.74	NA	NA
CF	0.44	0.2	-0.19 – 0.19	NA	NA
ISO. FLEX.	0.86	<0.05	(0.50 – 0.96)	6.70 (24.36%)	18.54 (6.46%)
T1/T3	ICC (2,3)	p-value	CI	SEM	MDC
CE	0.36	<0.05	-0.16 – 0.74	NA	NA
ISO. EXT.	0.23	0.3	-0.26 – 0.77	NA	NA
CF	0.70	<0.05	-0.2 – 0.93	2.55 (13.52%)	7.05 (37.44%)
ISO. FLEX.	0.74	<0.04	-0.12 – 0.94	3.23(10.93%)	8.95 (30.26%)

Note: CE – Concentric extension; CF- Concentric flexion; Iso. – Isometric; Ext.: - Extension; Flex. – Flexion; T1 -Tester 1; T2 – Tester 2; T3 – Tester 3.

Inter-tester reliability for peak force data shows T1/T3 (the only with at least moderate reliability for both concentric and isometric tests) (see table 4.12) isometric testing produces a slightly smaller SEM(%) in knee flexion movements (13.52% and 10.93%).

4.6.2.2 *Angle of Peak Force*

As seen in table 4.13, inter-tester reliability for angle of peak force is poor for T1/T2, for knee extension (-0.10) and for knee flexion (0.13), whereas for T1 vs T3 the reported ICC for knee extension is 0.27 and knee flexion is 0.70 (SEM=21.06%).

Table 4. 13 ICC for angle of peak force comparison for T1/T2 and T1/ T3 (º)

T1/T2	ICC (2,3)	P-VALUE	CI	SEM	MDC
CE.	-0.10	0.6	-0.95 – 0.57	NA	NA
CF	0.13	0.4	-1.5 – 0.75	NA	NA
T1/T3	ICC (2,3)	p-value	CI	SEM	MDC
CE	0.27	0.3	-1.8 – 0.82	NA	NA
CF	0.70	0.03	-0.09 – 0.92	5.37 (21.06%)	14.88 (58.34%)

Note: CE – Concentric extension; CF- Concentric flexion; T1 -Tester 1; T2 – Tester 2; T3 – Tester 3.

4.6.3 Concurrent Validity

Results for concurrent validity analysis for peak force are presented next (table 4.14) firstly with Pearson´s correlations, followed by WLP regression analysis and finally Bland-Altman Analysis. As data showed low reliability for angle of peak torque, validity analysis was not performed. Table 4.14 shows moderate correlations between the IKD and ASSA v2 for both the isometric extension movements and for concentric extension movement tested at 30º/sec. The correlation was lower for the extension movement tested at 60º/sec. For flexion movements, the correlations are moderate to good, with a slightly lower value (0.65) for the comparison between the IKD 60º/sec and 30º/sec.

103

Table 4. 14 Pearson´s correlations for peak torque (IKD vs ASSA)

	CORR.	95% CI	P VALUE	DF	T
ISOMETRIC EXTENSION	0.55	0.18 – 0.79	0.006	21	3.03
IKD 30º/SEC EXTENSION	0.56	0.20 – 0.79	0.004	22	3.18
IKD 60º/SEC EXTENSION	0.44	0.05 – 0.72	0.03	22	2.31
ISOMETRIC FLEXION	0.75	0.49 – 0.89	3.423e-05	21	5.24
IKD 30º/SEC FLEXION	0.69	0.41 – 0.86	0.0001	22	4.58
IKD 60º/SEC FLEXION	0.65	0.34 – 0.84	0.0005	22	4.04

Results for WLP regression (table 4.15) show no proportional or fixed bias between the IKD tests and the new prototype, this is demonstrated if zero is present in the confidence intervals from the interception (*a*) – proportional bias – and presence of 1 on the confidence intervals from the slope (*b*). From the *b* slope, it is clear that Isometric extension is the only movement that slightly underestimates the results from IKD ($b = 0.945$) where all the remaining movements overestimate the gold standard. This is particularly visible in the isometric flexion test with ($b = 1.28$).

Table 4. 15 WLP regression analysis IKD vs ASSA for peak torque

MOVEMENT	A	CI (A)	B	CI (B)
IE – ISO IKD	4.16	-66.41 – 74.73	0.945	0.53 – 1.36
CE – IKD30º/SEC	-22.42	-92.83 – 47.99	1.05	0.61- 1.49
CE – IKD60º/SEC	-16.55	-87.89 – 54.79	1.10	0.62 – 1.58
IF – ISO IKD	2.68	-24.01 – 29.38	1.28	0.830 – 1.72
CF – IKD30º/SEC	1.17	-19.71 – 22.05	1.07	0.67 – 1.47
CF – IKD60º/SEC	-1.87	-24.85 – 21.11	1.28	0.78 – 1.78

Note: A – Interception; B – Slope; IE- Isometric extension; CE- Concentric extension; IF- Isometric flexion; CF- Concentric flexion

In table 4.16, the correlation between IKD and ASSA for the angle peak torque is presented, with low levels of correlations presented. For this reason, no further validity analysis was performed.

Table 4. 16 Correlation between IKD and ASSA – Angle of peak torque

	CORR.	95% CI	P VALUE	DF	T
IKD30º/SEC EXTENSION	0.20	-0.22-0.56	0.34	22	0.97
IKD60º/SEC EXTENSION	0.29	-0.13-0.62	0.17	22	1.42
IKD30º/SEC FLEXION	0.003	-0.4-0.41	0.99	22	0.01
IKD30º/SEC FLEXION	0.25	-0.18-0.60	0.23	22	1.22

Note: Corr. - Correlation

Presented in the following sections are the tables and graphs describing the agreement between ASSA v2 and the IKD. Comparisons were assessed between isometric movements from ASSA and the IKD. At the same time, ASSA v2 concentric tests were compared with two different IKD speeds (30º/sec and 60º/sec). Log transform graphs were reported for all tests where the log-transform did not eliminate bias when calculating the differences.

4.6.3.1 *Knee extension*

Log transform Bland-Altman analysis did not remove bias (figure 4.8) therefore, Bland-Altman analysis (ratio by means) was performed. For knee extension, when comparing isometric tests, ASSA v2 values (figure 4.9 and table 4.17) were between -38% and 51% of those from the gold standard while revealing an average underestimate of -2%.

Figure 4. 8 Bland-Altman plot isometric knee extension ASSA vs IKD (log transform)
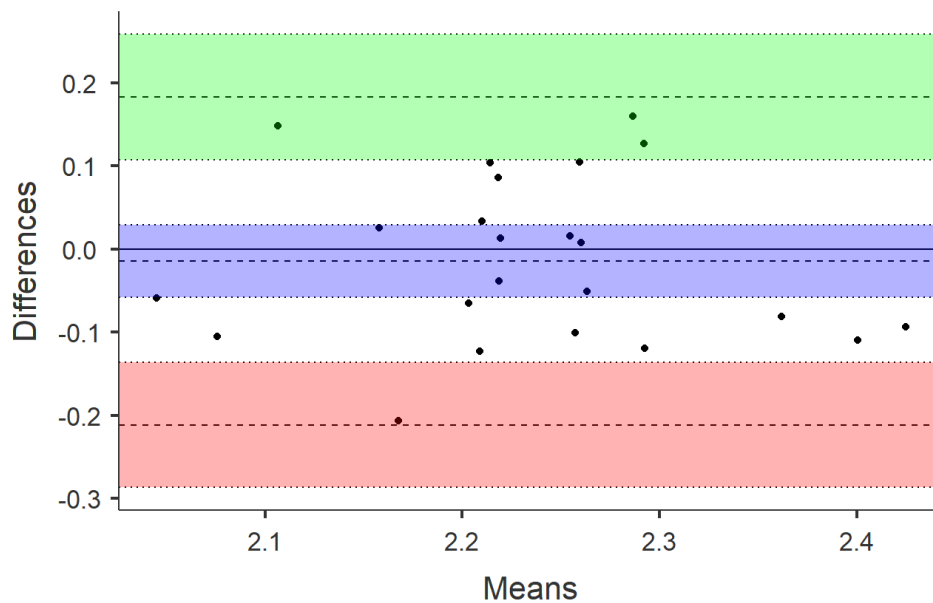
Figure 4. 9 Bland-Altman plot Isometric knee extension ASSA vs IKD in Nm
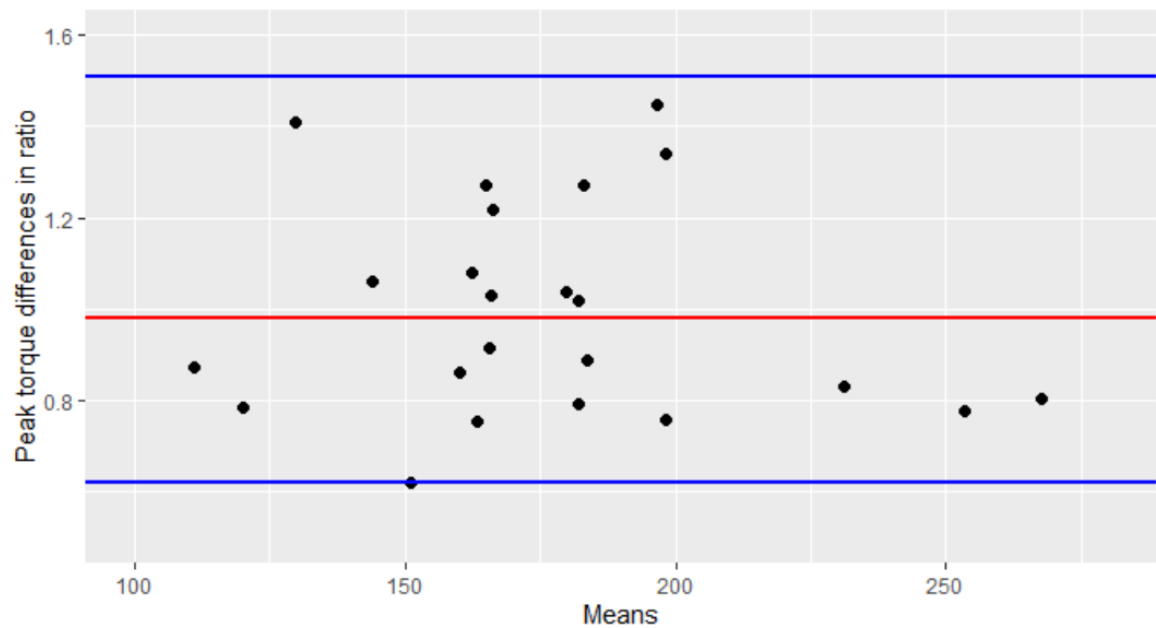
Table 4. 17 Bland-Altman LOA Isometric extension– ASSA vs IKD in Nm (ratio and percentage)

ISOMETRIC EXTENSION	ESTIMATE (RATIO (%))	LOWER	UPPER
BIAS (N = 23)	0.98 (-2%)	0.89 (-11%)	1.07 (7%)
LOWER LOA	0.62 (-38%)	0.52 (-48%)	0.72 (-28%)
UPPER LOA	1.51 (51%)	1.26 (26%)	1.82 (82%)

Results from knee extension concentric test (ASSA vs IKD 30º/sec – figure 4.10) demonstrate a negative bias of -13.80Nm with LOA from -82.19 to 54.59; table 4.18 and figure 4.11 display a negative bias of -7% with 95% of the results from ASSA fitting between -44% and 30% of the gold-standard.

Figure 4. 10 Bland-Altman plot concentric knee extension - ASSA vs IKD 30º/sec in Nm
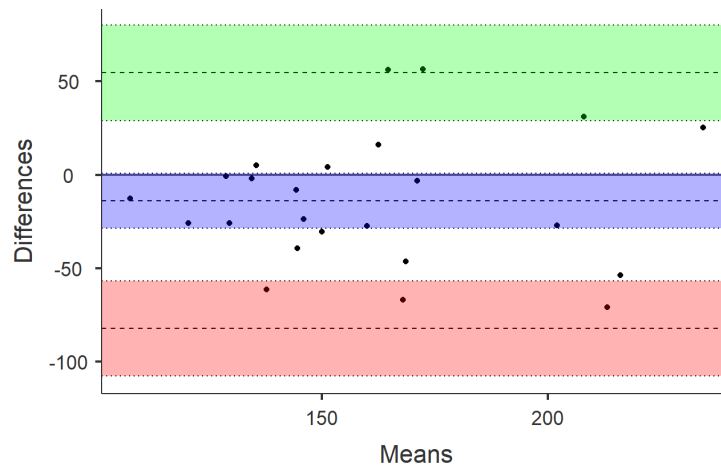
Figure 4. 11 ASSA concentric knee extension vs IKD 30º/sec in Nm
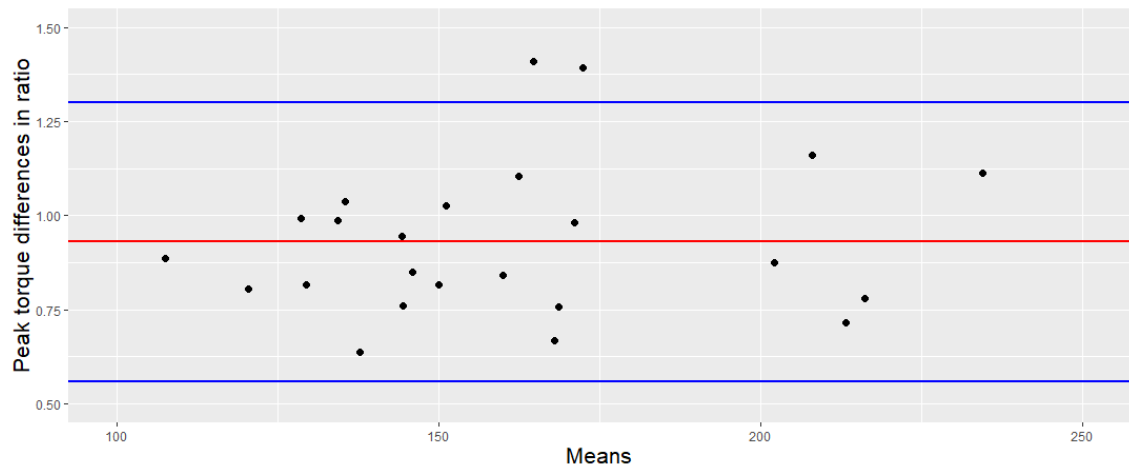
Table 4. 18 Bland-Altman LOA concentric knee extension - ASSA vs IKD30º/sec in Nm (ratio and percentage)

IKD 30º/SEC	ESTIMATE	RATIO	LOWER	UPPER
BIAS (N = 24)	-13.80	0.93 (- 7%)	-28.53	0.93
LOWER LOA	-82.19	0.56 (-44%)	-107.72	-56.65
UPPER LOA	54.59	1.30 (30%)	29.05	80.12

BA plot from the comparisons between ASSA concentric extension and the IKD 60°/sec (figure 4.12) reveals a small but positive bias, where ASSA appears to overestimate the IKD 60°/sec (figure 4.13) results by 1% lower LOA -44% and the upper LOA 46% (table 4.19).

Figure 4. 12 ASSA concentric knee extension vs IKD 60°/sec in Nm
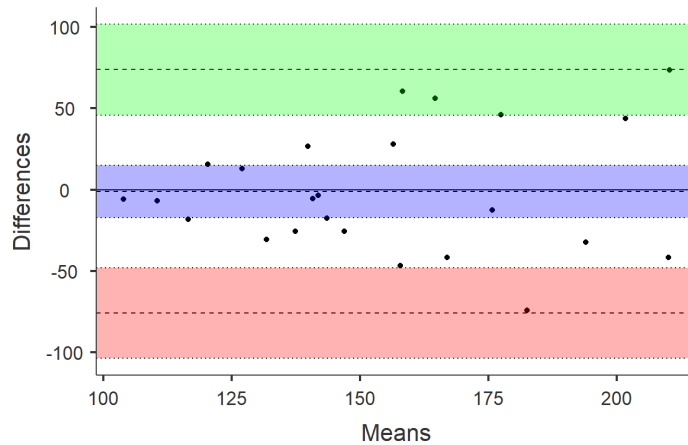
Figure 4. 13 Concentric knee extension ASSA vs IKD 60°/sec in Nm
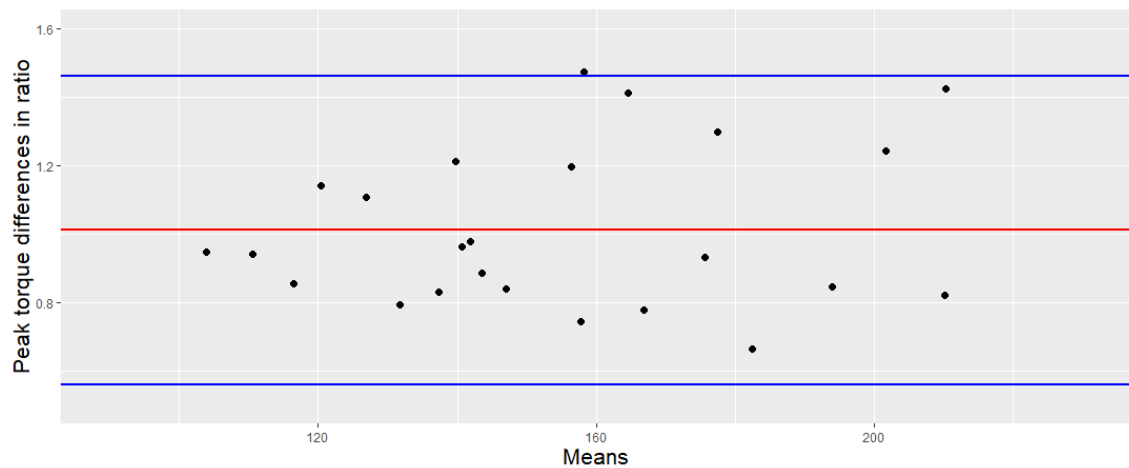
Table 4. 19 Bland-Altman LOA concentric knee extension - ASSA vs IKD60º/sec in Nm (ratio and percentage)

IKD 60º/SEC	ESTIMATE	RATIO (%)	LOWER	UPPER
BIAS (N = 24)	-0.98	1.01 (1%)	-17.09	15.12
LOWER LOA	-75.76	-0.56 (-44%)	-103.68	-47.83
UPPER LOA	73.79	1.46 (46%)	45.87	101.71

4.6.3.2 **Knee flexion**

The BA analysis using log-transform for isometric flexion did not eliminate bias and therefore a V-shaped 95% confidence interval was created (figure 4.14), this can be done using the BA method or using the one suggested by Ludbrook (2010) via WLP regression. As the WLP data was already available from previous analysis, the Ludbrook (2010) method was chosen. Isometric flexion data (figure 4.15 and table 4.20) shows an overestimate of 32% by ASSA, while 95% confidence intervals place the LOA between 104% and -16% of the IKD.

Figure 4. 14 V-shaped LOA Isometric flexion – ASSA vs IKD in Nm
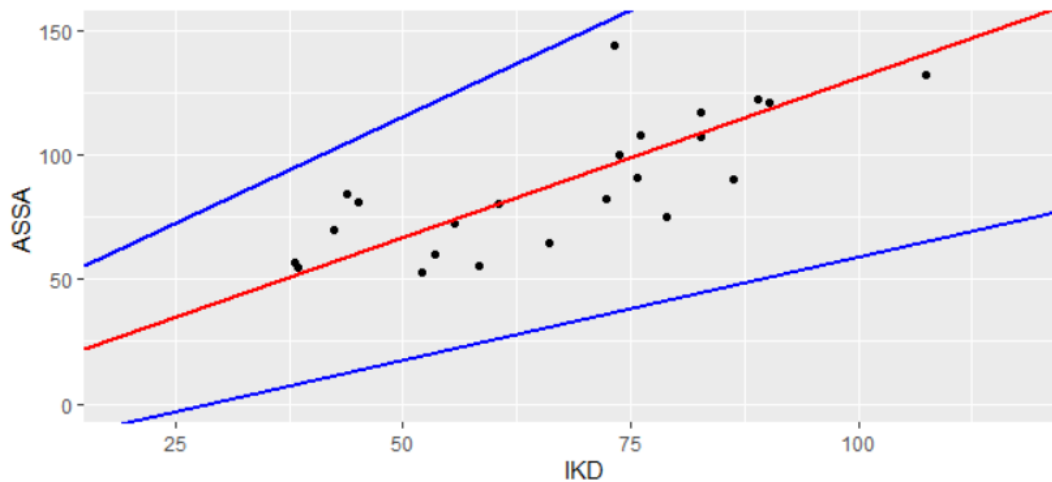
Figure 4. 15 Bland-Altman plot knee flexion isometric - ASSA vs IKD in Nm
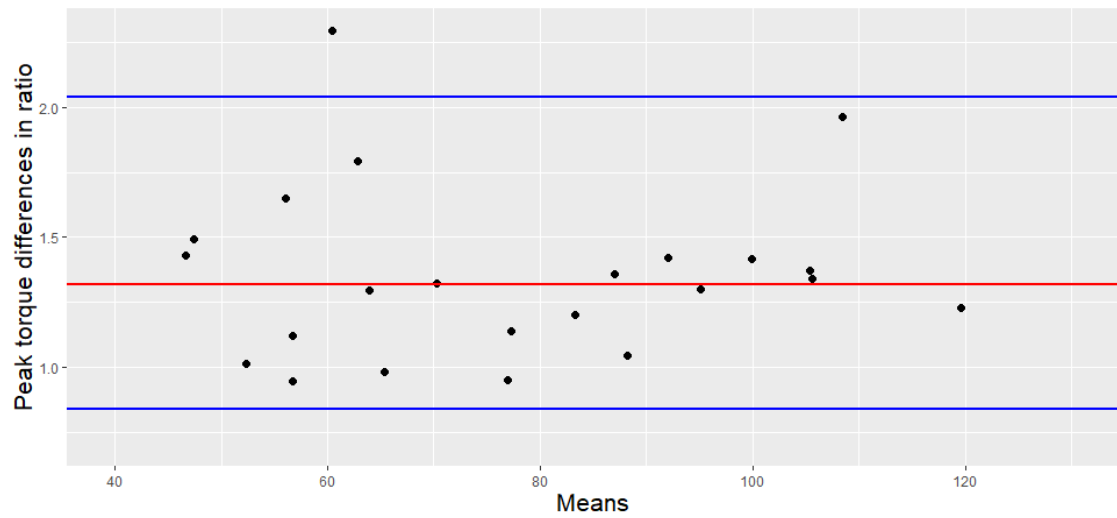
Table 4. 20 Bland-Altman LOA Isometric knee flexion – ASSA vs IKD isometric in Nm (ratio and percentage)

ISOMETRIC FLEXION	ESTIMATE (RATIO (%))	LOWER	UPPER
BIAS (N = 23)	1.32 (32%)	1.19 (19%)	1.45 (45%)
LOWER LOA	0.84 (-16%)	0.71 (-29%)	1.0 (0%)
UPPER LOA	2.04 (104%)	1.74 (74%)	2.44 (144%)

The comparison between concentric knee flexion and IKD 30º/sec reveal a slight bias of 5.04Nm (12%) from the prototype (figure 4.16), while ratio analysis in percentage shows the LOA are between -37% and 61% (figure 4.17 and table 4.21).

Figure 4. 16 Bland-Altman plot concentric knee flexion ASSA vs IKD 30º/sec in Nm
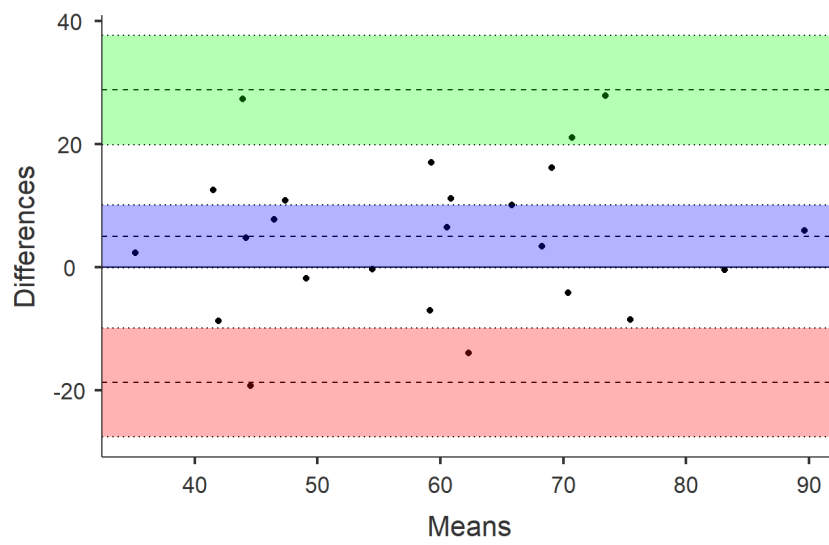
Figure 4. 17 ASSA concentric knee flexion vs IKD30º/sec in Nm
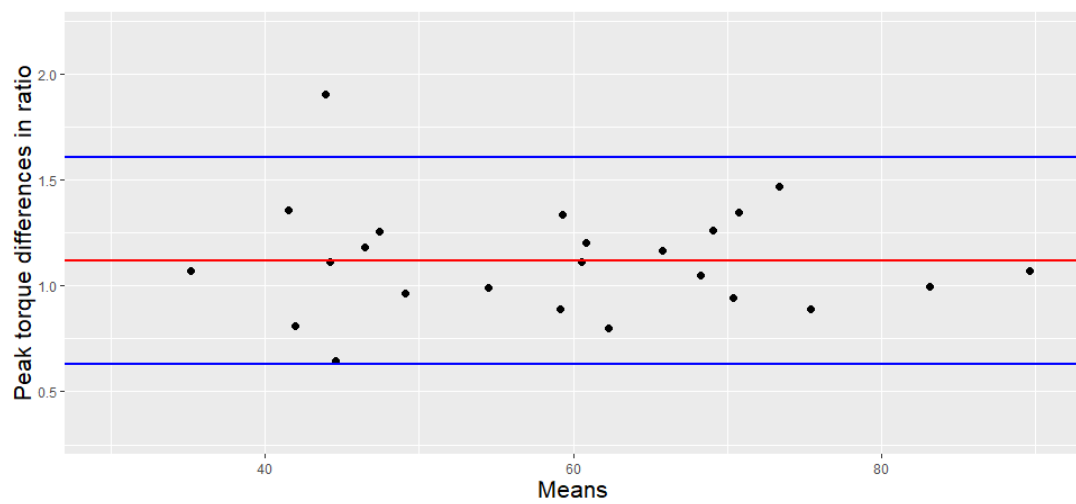
Table 4. 21 Bland-Altman LOA knee flexion - ASSA vs IKD 30º/sec in Nm (ratio and percentage)

	ESTIMATE	RATIO (%)	LOWER	UPPER
BIAS (N = 24)	5.04	1.12 (12%)	-0.08	10.16
LOWER LOA	-18.73	0.63 (-37%)	-27.61	-9.86
UPPER LOA	28.82	1.61 (61%)	19.94	37.69

BA analysis for ASSA concentric knee flexion vs IKD 60º/sec (figure 4.18) demonstrate ASSA overestimates the IKD, by an average of 24% (figure 4.19 and table 4.22), with an lower LOA of -21% and a upper LOA of 95%.

Figure 4. 18 Bland-Altman plot concentric knee flexion ASSA IKD 60º/sec (log transform)
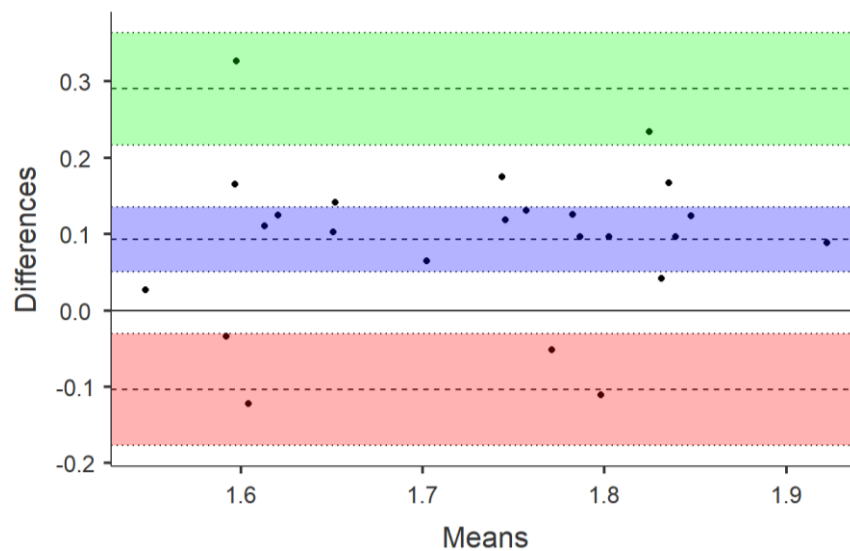
Figure 4. 19 Bland-Altman plot concentric knee flexion ASSA vs IKD 60º/sec in Nm
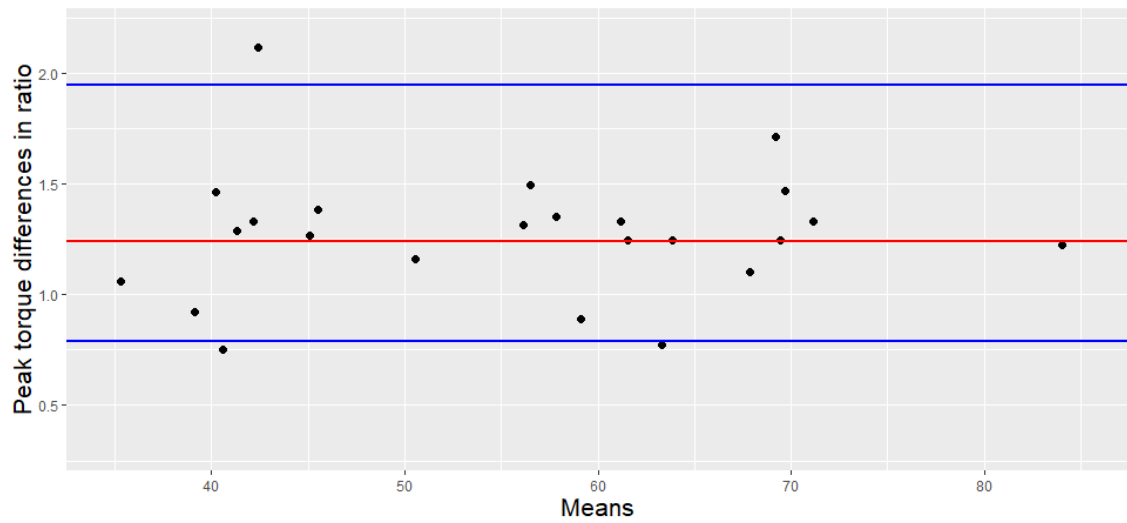
Table 4. 22 Concentric knee flexion vs IKD 60º/sec (ratio and percentage)

ISOMETRIC EXTENSION	ESTIMATE (RATIO (%))	LOWER	UPPER
BIAS (N = 24)	1.24 (24%)	1.12	1.35
LOWER LOA	0.79 (-21%)	0.68	0.93
UPPER LOA	1.95 (95%)	1.62	2.29

4.7 Discussion

Physiotherapists assess strength regularly in their everyday practice, and most professionals use manual muscle tests. Evident problems with manual muscle tests have been identified (see chapter 3, section 3.2.4) - nonetheless, these are still the dominant approach in clinical practice. The current pilot study was critical on this prototype's early-stage development by investigating clinimetric properties to establish intra and inter tester reliability and validity. The study also contributed to establishing an adequate approach for future trials.

The pilot study found that smaller peak force values were retrieved when the concentric test is used for knee extension and flexion – compared to the isometric test. This was already a documented fact from previous research comparing isometric and concentric muscle actions (Doss and Karpovich, 1965; Knapik, et al., 1983) but to the author's knowledge has never been demonstrated in manual muscle tests while using a concentric HHD. The overall strength difference might be an essential point favouring the concentric approach mainly when testing stronger patients or at a later stage in rehabilitation where most physiotherapists might struggle to get reliable results due to the participants' strength.

4.7.1   Intra-tester Reliability and responsiveness

4.7.1.1   Between-session peak force

The results for intra-tester reliability (between-session ICC(2,3)) showed that tester 1 had good reliability for average peak force in all movements (ICC>0.75). Concentric testing was at least as reliable as isometric testing and more reliable in the knee extension movement. Moreover, SEM for concentric tests for knee extension and flexion is also smaller (5.04% - 8.9%) when compared with isometric testing (5.9% - 9.8%), respectively. MDC results also suggest that concentric testing (extension: 5.88kg (14%); flexion: 4.99kg (24.74%)) appears to be superior – by detecting smaller differences - when compared with isometric testing (extension: 9.65kg (20.4%); flexion - 7.54kg (27.2%)). These results are statistically significant (p<0.01), as demonstrated in table 4.6 and are also useful for future reference in similar research for concentric HHDs as they have never been reported before. A word of caution is also important as these results might not be identical for all other HHDs on the market.

4.7.1.2   Between-session angle of peak force

In contrast, for the average angle of peak force, the reliability between sessions was moderate for knee extension (ICC 0.61) and for knee flexion (ICC 0.71). SEM(%) stood at 9.28% for knee extension with MDC(%) at 25.70%, meaning the test can detect changes in peak force angle if they are superior to 8.34°, whereas for knee flexion, these

are slightly larger with 18.52% (SEM) and 51.29%(MDC). Le-Ngoc and Janssen (2012) reported similar values for their new prototype in a knee extension movement (ICC = 0.75 CI:0.38-0.91). However, the CI (-0.001 – 0.85) for the current prototype is wider, which raises questions about its clinical value and needs to be improved further by adapting its algorithm. One of the main reasons for the discrepancy may be software differences revealed by Le-Ngoc and Janssen (2012) as they explained that their device did not need to be constantly aligned with the segment being tested for ROM readings. This is not the case with ASSA.

There are no previous data published for the concentric angle of peak force using an HHD for knee flexion, but the results from this pilot study are encouraging as the ICC (0.71) is superior to angle of peak force for knee extension, and this suggests that different joints or testing positions might offer acceptable results.

For comparison of the reliability between the gold standard and ASSA, data from three repetitions for the angle of peak torque from the IKD was collected (ICC(2,1)). The ICC from this study and from Le-Ngoc and Janssen (2012) regarding the IKD ability to detect angle of peak torque was smaller than other research. This might rise from the short training procedure with the subjects. This is evident when comparing both results with the data from Maffiuletti, et al. (2007) where ICC values are larger for both knee extension and knee flexion angle of peak torque but where researchers allowed for 20 repetitions for each movement before testing began as a form of warm-up.

4.7.1.3   Between repetitions peak force

Data comparing the reliability between different peak force values between three repetitions (ICC(2,1)) for each of the testers showed good reliability for both tester 1 and tester 2 in each of the tested movements (concentric flexion and extension), whereas tester 3 displayed poor reliability with both concentric movements displaying ICC values of less than 0.5. This might be due to the possibility that tester 3 tested participants who produced larger torques and therefore were inherently more difficult to test. This should be investigated further in the next chapter by using testers with more experience.

4.7.1.4   Between repetitions angle of peak force

The angle of peak force data demonstrated fair to moderate reliability for knee flexion movements for tester 1 and tester 3 (>0.55) whereas tester 2 has poor reliability (<0.36). Knee extension reliability for angle of peak force is poor for tester 1 (day 1 and day 2) and moderate for tester 2 (0.53) and tester 3 (0.55). Considering knee extension also produced the higher peak torque for the trial and tester 1 resisted the higher forces in that movement, it might indicate that it is more challenging to maintain the position of the device across the range of motion at higher values of the force. This seems to be supported by an increase in reliability for testing knee flexion by tester 1 (0.59 – day 1; 0.68 – day 2) and for tester 3 (0.59) but not for tester 2 (0.43). Further data needs to be collected to draw further conclusions.

4.7.2   Inter-tester Reliability and responsiveness

4.7.2.1   Peak force

Inter-rater reliability for peak force results were divided into two pairs T1/T2 and T1/T3. For the pairing T1/T2, reliability for peak torque was poor in knee extension isometric (0.32), concentric (0.18) and knee flexion concentric (0.44) with the only acceptable value being the test for isometric knee flexion (0.86). When looking at T1/T3, reliability was also poor for knee extension isometric and concentric (0.23-0.26), but higher reliability was shown on knee flexion isometric and concentric (0.74-0.70). The main explanation for this finding is that testers 2 and 3 did not possess any MMT training, which could have impaired the comparison with an experienced tester. Further analysis is needed in the next chapters to improve insight.

4.7.2.2   Angle of peak force

Reliability for inter tester angle of peak force is poor for all movements except for T1/T3 knee flexion, even though CI intervals for this movement were quite wide, which shades doubt over its clinical acceptability. Inter-tester reliability for knee extension is poor for the two testing pairs and needs to be improved if it is to be useful for clinicians.

Several reasons arise that could contribute to the current results for both (inter-rater reliability) peak torque and peak torque angle. Firstly, each tester needed to assess at least 17 participants for a statistically significant ICC result that was not achieved due to a tester's inability to terminate the data collection. Therefore, ICC values for inter-rater reliability should be interpreted with this in mind, and further research is needed to clarify this point. Secondly, the testers were not able to resist several movements, as reported in section 4.6, which might be due to their inexperience and the device's design. The device's length can create instability while testing, and the handling area might affect comfort. These factors might increase the difficulty in performing an adequate test and influence the results. Apart from these issues, low-reliability scores might be related to the current testing positions, since the knee extension testing position requires more skill due to decreased stability both from the tester and participant. It may also be hypothesised that concentric testing requires more training than isometric, particularly for inexperienced testers. These issues were further investigated in the next stage of the thesis.

4.7.3   Validity

The correlation obtained suggests a moderate degree of association between the devices for knee extension tests in terms of validity. Whereas for knee flexion tests, the correlation was good, showing a superior association with the isometric test than the concentric tests. Correlations were superior when testing at lower speeds (30º/sec), this is an important point as in future trials only one testing speed will be assessed. In general terms, the correlation values are inferior to similar trials (Hansen, et al., 2015; Martins, et al., 2017), particularly in knee extension, and should be improved to superior levels to increase the validity of ASSA compared with the gold standard.

Validity test results for the WLP regression analysis (see table 4.15) show peak torque is valid as it does not show fixed proportional bias for the isometric and concentric tests in both knee extension and knee flexion. From the same table, the slope values indicate under or overestimation for IKD comparison (if the values were 1, it would indicate the values from ASSA are perfectly matching the IKD) are closer to 1 when comparing with the IKD at 30º/sec and therefore for the next stage of testing the comparison will be done using that speed instead of two different speeds such as in this pilot.

The LOA presented are smaller for the concentric test using the current prototype when comparing the isometric and concentric movement for knee extension, which suggests that the concentric approach might be better at conveying valid data than the gold standard. Previous research by Martins, et al, (2017) for belt stabilised HHD knee extension was 40% more or -66% less than the average force displayed; these values are slightly better than the pilot study results for knee flexion but their HHD was stabilised. For the comparison with the IKD30º/sec vs HHD, LOA are lower for the concentric approach displayed here (LOA -44% to 30%). These are positive signals favouring the concentric approach as agreement seems to be narrower with this approach.

For knee flexion tests, isometric knee flexion LOA stand at -16% to 104%, whereas previous research by Martins, et al., (2017) has reported LOA for IKD vs HHDs at 60.57% and -42.74%. In the concentric knee flexion the LOA range from 61% to -37% (IKD 30º/sec) and 95% to -21% (IKD 60º/sec).

This means the device is valid as it conveys the same construct; however, the output cannot be used interchangeably with the IKD and cannot replace the IKD. When comparing the current data with similar research (see Chapter 8), there is limited research using the LOA for validity testing as most authors are still using ICC and other correlations. The results from the concentric approach convey narrower LOA than isometric tests. Therefore, it can be considered at least as appropriate for clinical use as fixed HHD (further details can be found in chapter 8).

The poor results (see table 4.16) for the correlations in peak torque angle suggest that the prototype cannot convey valid results of peak torque angle. Consequently, WLP regression and BA analysis were not performed. Further improvements are needed to be able to use this feature in a clinical setting, and more data regarding this issue will be made available in the next chapters.

4.7.4 Strengths and limitations

A strong point in this chapter was the good intra-tester reliability and the validity found. From the pilot test, it is also possible to ascertain that the prototype was able to perform

as expected. Considering this is a newly developed prototype, the results support the development of a new version and further testing with specialised testers.

One of the reasons for the small to moderate correlations regarding peak torque and angle of peak torque might be twofold. Firstly, the length of the device might contribute to some instability due to an increased lever when compared to common HHD, this might make it more difficult to resist the concentric movement and to maintain a similar speed through the range of motion. Secondly, the current software might also contribute to this type of error. Currently, the device provides an angle output in a consequential form, not allowing for repeated angle values – this might prevent the device to get accurate readings if the segment is moving too fast, too slow, or in an uncontrolled way. This might justify why tester 1 (who showed the biggest forces) had lower correlation values in knee extension movement when compared to the other testers.

Another limitation was the absence of segment size measurements; this prevented the direct comparison with data from the IKD, which meant that a regression procedure had to be run to estimate each participant's segment length. This might have biased the results for the validity assessment as such formula has an estimated error of around 2-3cm. This issue was addressed on the next stage of data collection by collecting segment size data.

As a take from the pilot study, the slower speed (IKD 30°/sec) showed a narrower width for the LOA; therefore, this will be the selected speed of analysis for the research's next steps. Further testing is also needed to gather data for inter-tester reliability using concentric tests, particularly with an experienced and homogenous group of testers.

4.8 Conclusions

These are encouraging results to try and use the newly tested approach and device as it indicates that the device is valid regarding peak torque output when compared to the gold-standard IKD. The concentric movement technique demonstrates intra-tester reliability and it seems to provide more consistent results for peak torque than isometric testing.

5. INTRA-TESTER, INTER-TESTER RELIABILITY AND VALIDITY OF A NEW HHD IN A GROUP OF EXPERIENCED PHYSIOTHERAPISTS

5.1 Introduction

The current chapter describes the testing of ASSA v3. As stated in chapter 3, ASSA v3 is an upgraded version of the prototype tested in the previous chapter that has been enhanced (in terms of software and hardware). With data from the pilot study an improved and more user-friendly version of the prototype was developed. Previously, we have investigated how an early-stage prototype could be employed in manual muscle testing by physiotherapists. In order to achieve this, the HHD was used on several joints with a larger, heterogeneous and experienced group of physiotherapists, thus exploring its on performance.

5.2 Objective

The current investigation's objective was to test ASSA v3 validity and reliability in several joints with experienced physiotherapists. This was done by comparing concentric manual tests on shoulder abduction, elbow flexion, knee extension and flexion comparing movements between testers and with an IKD. Four different physiotherapists performed the testing in different sessions with a healthy adult population.

5.3 Methods and Materials

5.3.1 Advancements in the current prototype

The first step in the development of ASSA v3 was to enhance its shape so that we could improve the product's usability. With that in mind, all components were enclosed in a new casing (15cm x 8cm x 5cm) that was then adjoined to a wood attachment 3.5cm in height and that would be directly in contact with the participant. This decreased the height of the device by more than 40%, allowing for increased control which should aid

reliability on both angle and force detection. A new, lighter load cell (beam load cell, capacity: 0-120kg) was essential in decreasing the weight of the device while maintaining its accuracy. The software was also updated with new modes of data collection and output. ASSA v3 provides complete (raw) data output – the previous version filtered data before producing the excel file – so it can be fully analysed afterwards (Figure 5.1).

Figure 5. 1 Excel output example from ASSA v3

	A	B	C	D	E	F
1	raw data					
2	time	force (kilos)	angle			
3	0.09863	1.69	0.33104			
4	0.19531	1.71	0.27543			
5	0.29102	1.75	0.44411			
6	0.38867	1.87	0.99256			
7	0.48633	2.22	2.21069			
8	0.58203	2.97	3.68406			
9	0.67871	3.81	4.57639			
10	0.77539	4.62	7.68579			
11	0.87305	5.52	8.42746			
12	0.96973	6.36	8.9196			
13	1.06543	6.97	11.3118			
14	1.16211	7.5	12.1215			
15	1.26074	7.81	15.6244			
16	1.35645	8.23	16.1559			
17	1.45313	8.72	19.162			
18	1.5498	9.38	16.6448			

Sheet1 Sheet2 Sheet3

Time information was displayed at the fastest rate possible by the load cell amplifier while force data was not obstructed by the software, which was previously filtering maximum force per angle. In practical terms, the device (ASSA v2) would only register information on a unidirectional fluid movement. This new prototype can detect errors or sudden movements as it allows a freer-flowing data stream. The device was calibrated daily before data collection to ensure stable and accurate measures. This was performed using the repeatability method, the HHD was turned on and zeroed, then a load of 50% of the load cell capacity was positioned on the measuring area and the result collected. This was repeated five times, if the results were off by more than ±0.05% a re-calibration was performed.

5.3.2   Design

A single-blinded repeated measures RCT to test reliability and concurrent validity was done. One day was reserved for collecting data with the IKD, and three days for the HHD. For the HHD, tester 1 and 2 performed two days of assessment and tester 3 and 4 only one day. Movements tested were shoulder abduction; elbow flexion; knee extension and knee flexion. In order to minimise fatigue participants had a minimum of 48h between testing and were tested at similar times of day. Neither participants nor testers could visualise the results for each repetition. Each intervention's order was randomised (as well as the tester who assessed the participant first) to minimise bias, but the order of movements executed was maintained to facilitate the procedure.

Data from all testers (1, 2, 3, 4) was used to allow for inter-tester reliability analysis. Validity testing was performed on an IKD (Biodex Multi-Joint System 4; Biodex Medical Systems, Inc) at 30º/sec (see section 4.5.3). Researchers present on the day provided consistent information to participants to provide uniform information and verbal incentive, a script was developed for both the IKD and HHD.

Torque calculations were performed to allow for comparison with IKD data. Moment arm was determined by measuring the segment size with a soft measuring tape in each participant. This method has proved to be a reliable way to assess leg length (Beattie, et al., 1990). Segment size was measured considering the perpendicular distance from the axis of rotation to the dynamometer's placement point. Shoulder segment was measured from the lateral aspect of the acromion to the posterior base of the first metacarpophalangeal joint; elbow segment was determined from the lateral epicondyle to the anterior projection of the radial styloid process; lower leg segment was determined from the centre of the lateral condyle of the femur to the centre of the lateral malleoli. To allow for a good contact area with the device, resistance was applied to a pen marked dot 3cm above the anatomical points described above – this dot was considered the central point for HHD contact and was marked again for the second session. The prototype was maintained perpendicular – as much as possible - to the segment through the range of motion.

5.3.3 Recruitment and participants

Participants were healthy adults recruited from the Higher School of Health of The Portuguese Red Cross (HSHPRC). They all received written and verbal information about the study and signed a consent form before the study began. The trial had Ethical approval from ARU (code: ESPGR-05) (Appendix 8) as well as HSHPRC where the data was collected.

To determine the sample size both reliability and validity were considered for the analysis. Regarding reliability, to achieve a power of 90% and alpha of 0.05, a difference of acceptable reliability (0.70) – which we determined in the previous data collection - to expectable reliability (0.90) this study would need to recruit at least 18 individuals with three observations per participant (Bujang and Baharum, 2017). Regarding validity we applied Bland and Altman formula (1.71SD/sqr(n)) for the calculation of the standard error (SE). From data collected in the pilot study and for a power of 90% and alpha of 0.05 and a maximum acceptable difference between methods of 120Nm (for knee extension), the minimum required sample size was determined to be 17. Since some data might be lost and considering that participants might withdraw, it was deemed more appropriate to recruit 20%-30% more participants. A total of 24 subjects were recruited.

Participants were included if they hadn't had surgery or serious injury in the last six months or any current injury on back, lower/upper limbs. We excluded anyone if they were pregnant; had any injury that would prevent them from exerting maximal force on any of the joints to be tested; blood pressure above 199/109 mmHg; previous cerebral aneurysm; cerebral haemorrhage in the last six months; cataract surgery in the last six weeks; non treated inguinal hernia.

In terms of physiotherapists' recruitment to test the participants Koo and Li (2016), suggest at least three. Considering that data loss or dropout might occur, it was deemed that four testers would be appropriate.

Physiotherapists recruitment was based on judgement sampling from the available physiotherapists from the HSHPRC. Experienced physiotherapists inclusion criteria: more than ten years clinical experience; ability to perform a manual muscle test against strong resistance. Exclusion criteria: hearing deficit; inability to understand the protocol,

inability to understand English (both written/spoken). Physiotherapists performed no proficiency test as in a real-life scenario, these would not be performed. The recruited physiotherapists had between 23 and 13 years of experience.

Before the study began, testers underwent a 2h training to become proficient in the use of techniques in all joints. Intra tester reliability (between days) was assessed by tester 1 and tester 2 who evaluated participants on two different sessions (day 1 and day 2).

Once recruitment was finished and data collection started not all participants attended the planned sessions. On the day 1, tester 1 and tester 2 assessed all but one participant, whereas, on day 2, 21 participants were tested. Regarding tester 3, 22 participants were assessed and tester 4 assessed 21 participants. IKD data were collected on 19 participants.

5.3.4   Testing procedure

Participants were asked not to perform any vigorous physical activity 48h before each testing session and to maintain their nutritional and activity habits. Subjects were, as much as possible, scheduled to perform the test with at least one-week interval to warrant independence between trials as recommended by (Terwee, et al., 2007) and at similar times of the day to minimise any diurnal influence in strength production. Participants dominant side was determined beforehand by asking the participant which leg they used to kick a ball, and only that side was tested.

Participants performed a warm-up before testing started (in both the IKD and manual muscle test) where 1-3 repetitions was permitted (25-50% MVC) to allow for a gentle increase in muscle activity and task recognition before the MVC. In the literature, methods of MMT that use a dynamometer were developed considering an isometric contraction. For this study, however, the author adapted those techniques to match the IKD testing movements as closely as possible, as the testing procedure is not in a static position but through a certain range of motion. However, some of these positions might not make sense in a clinical setting – for instance, the shoulder abduction test where the resistance was applied distally on the hand to simulate the IKD position. As documented previously, even for regular HHDs using isometric testing, different researchers used

various approaches (Stark, et al., 2011). In this case, testing positions were selected to match as closely as possible the concentric testing on the IKD.

Below is presented the description for the adopted positions during testing when using the HHD (table 5.1 – where no straps or fixations were applied to the participants) and the IKD (table 5.2). All dots that marked the segment size were repeated on each day of testing and were used by all testers to position the dynamometer's centre at every single session. The verbal incentive was provided as soon as the test began with different inputs for each joint: shoulder – "lift your hand, up, up, up"; elbow – "bend your elbow, bend, bend"; knee extension -" extend your knee, up, up"; knee flexion – "bend your knee, bend, bend". Participants were also instructed to keep their elbow straight (on shoulder abduction), the wrist in neutral position (for elbow flexion) or the thigh resting (for knee extension and flexion) on the plinth as necessary. If the participant was struggling to follow any instructions, the test was repeated. Participants were given 3-5 seconds to rest between contractions, 1 minute between movements and 5-10 minutes between devices/testers to avoid fatigue and for repositioning (Whinton, et al., 2018).

Table 5.1 Hand-held dynamometer testing position

ACTIVITY	POSITION	ASSA
SHOULDER ABDUCTION	Participants sat on a stool next to a wall to prevent trunk inclination. Both feet were placed on the floor for increased stability and the arm not being tested - resting by the side of the trunk - the opposite shoulder resting on the lap or across the chest. Participants were explicitly asked not to push with the elbow (of the resting arm) against the wall.	The tester directed ASSA's contact area to the region between the 1st and 2nd metacarpal bones while maintaining a perpendicular orientation with the forearm, maintained throughout the movement. Participants were instructed to keep the elbow extended. The tester would stand facing the participant and holding the device with one hand.
ELBOW FLEXION	Participants sat, the arm to be tested resting on a plinth, which was tilted so the whole arm lays at around 45° with the floor. This simulates the testing position in the IKD.	ASSA's contact area on the forearm's anterior distal area using the mark done before as a reference. The tester was asked to maintain a perpendicular orientation throughout the movement.
KNEE EXTENSION	Participants sat on the edge of a plinth with back supported against wall (hard foam pillows supplied if needed) so that the knee sits outside the plinth. Knee relaxed, hands on plinth for support. Not allowed to hold plinth to extend knee.	The tester directed ASSA's contact area to the anterior lower third of the tibia just above the tibial malleoli while maintaining a perpendicular orientation with the lower leg segment, maintained throughout the movement. Tester faced the subject

		while the participant extended the knee (maintaining upper thigh in contact with plinth at all times).
KNEE FLEXION	Participants sat on the edge of a plinth with back supported against wall (hard foam pillows supplied if needed) so that the knee was kept outside the plinth. Knee fully extended, hands on plinth for support. Not allowed to hold plinth to extend knee. Participant to avoid pressure (or to "rest") on prototype before the tester initiates the test.	The tester directed ASSA's contact area to the posterior lower third of the tibia just above the tibial malleoli while maintaining a perpendicular orientation with the lower leg segment, maintained throughout the movement. Tester faced the subject. Patient flexed knee while maintaining upper thigh in contact with plinth at all times.

Figure 5.2 ASSA shoulder abduction

Figure 5.3 ASSA Elbow flexion

Figure 5.4 ASSA Knee extension and knee flexion

Figure 5.5 IKD Elbow flexion and IKD Shoulder abduction

Figure 5.6 IKD Knee extension and knee flexion

Table 5.2 IKD position

ACTIVITY	POSITION
SHOULDER ABDUCTION	Subjects were asked to sit on the IKD chair, to cross the arms and not hold the testing chair during the procedure. Chair upright at 85°, shoulder attachment installed, and the dynamometer axis aligned with the acromioclavicular joint axis.
ELBOW FLEXION	Subjects sat on the IKD chair, unused arm across the chest. The chair was upright at 85°, and an upper arm attachment provided a 45° plane for the humerus to rest. An attachment was used so that the testers were holding an IKD handle, and the dynamometer axis was aligned with the lateral elbow epicondyle.
KNEE EXTENSION	Subjects were asked to sit in the IKD chair, to cross the arms and not hold the testing chair during the procedure. IKD axis was aligned with lateral femoral condyle. The chair was positioned so the participant could fully extend and flex the knee.
KNEE FLEXION	Same as above

5.4 Data processing

A customised RStudio (Version 1.2.5033 © 2009-2019) program was written to analyse the data collected from the ASSA and the IKD. Data processing consisted of the following steps:

1. A low-pass 4th order Butterworth filter was designed to filter the data from the devices. As the devices had different sampling rates (ASSA – 10Hz; IKD – 1000Hz), a cubic spline interpolation for the device with the lowest sample rate was used to allow for a more accurate comparison. Data were tested for normality using Shapiro-Wilks ($p < 0.05$).

The prototype captured strength in kilograms (kg). These values were later converted to torque in Newtons (Nm) to allow for comparison with the IKD - the following formula was applied (Martin, et al., 2006; Sung, Yi and Shin, 2019):

HHD output (kg) x 9.81 x lever length (m) = torque (Nm)

2. Gravitational moment correction for the ASSA force output (when comparing the IKD) was done using the method suggested by Kellis and Baltzopoulus (1996), the authors used anthropometric data from Dempster (1955). The knee joint gravitational correction was: $M = (l*0.437)*(0.06*BW)$, considering M = gravitational moment, l = length of the limb (m), 0.437= centre of mass proximal joint distance/segment length, 0.06 = leg-foot weight/body weight, and BW = body weight (N). The equation suggested by Kellis and Baltzopoulus (1996) was adapted, anthropometric data presented by Dempster (1955) as in the previous chapter. For shoulder abduction, centre of mass proximal joint distance/segment length = 0.512 and upper limb weight/body weight = 0.049. Whereas, for elbow flexion data, centre of mass proximal joint distance/segment length = 0.677 and forearm weight/body weight = 0.022. IKD and ASSA both produce joint angles in degrees (°) so no change was made to this parameter.

3. Angular impulse data was calculated using the area under each curve. It was calculated for reliability and validity. For the reliability analysis, the full range of movement was used for the calculation. Whereas for the validity analysis, to minimise discrepancies in range of movement between ASSA and the IKD, the following range of movement were analysed for each device. Elbow flexion: 5 – 95°; Knee flexion: 15 – 105°; Knee extension: 15 – 105°; Shoulder abduction: 5-95°. The range of movement for each joint was chosen considering the most common available range of movement for all testers taking into account the testing position. The range is different for knee movements as most patients never achieved less than 5 degrees of knee extension due to the testing position.

4. Mean and standard deviations (SDs) are reported. Due to the nature of the analysis where comparisons between day or tester were performed, if a data point was missing for a specific participant, then the correspondent data point for the following day/tester was also removed. Therefore, different numbers of participants are expected.

5.4.1   Data analysis

Intra-tester reliability and responsiveness: The degree of correlation as defined by Schrout and Fleiss (1979) was done using a mean of three repetitions and comparing

the data between different days for tester 1 and tester 2. All the peak force, angle of peak force and angular impulse tests from ASSA was calculated using ICC (2,3). The ICC (2,3), uses a two-way random-effects model, with mean scores. SEM and MDD in absolute values and in percentage were also obtained.

Inter-tester reliability and responsiveness: The approaches to analyse this included ICC, SEM, SEM in percentage, MDC and MDC in percentage . The ICC (2,3), reflects a two-way random effects model, using average scores from 3 repetitions.

SEM results were obtained by multiplying SD_1 the standard deviation from the first session output by the square root of one minus the ICC ($SEM = SD_1\sqrt{1 - ICC}$) this was then converted to SEM% (SEM% = (SEM/mean) × 100) (Portney and Watkins, 2015). MDC were calculated with: MDC = z × SEM × $\sqrt{2}$, z = 1.96 (for 95% confidence intervals), this was also reported in percentage (MDC% = (MDC/mean) × 100) (Lexell and Downham, 2005; Portney and Watkins, 2015). Similarly to the intra-tester reliability peak force, angle of peak force and angular impulse were calculated, to facilitate comparison with similar research on both inter and intratester reliability.

Concurrent Validity: To quantify the agreement between the two devices (ASSA vs IKD), Pearson's r correlation, Bland Altman analysis – Limits of Agreement was used (Zaki, et al., 2012) and WLP regression was used to assess correlation between the two devices and provide a calibration equation (Ludbrook, 2010). Data analysis followed these steps: Bland Altman analysis – Limits of Agreement was used to test for agreement between the two devices (IKD vs ASSA); 95% LOAs were calculated as LOA = bias ± 1.96 SD (Zaki, et al., 2012; Martins, et al., 2017). Shapiro-Wilk (p <0.05) test was used to assess normality. When the data set was determined to be non-normal (line of best fit different from zero), a log-transformed BA analysis was done. If data was not normally distributed a log transformation would be performed, if log transform bias was still present (heteroscedasticity) then a linear regression with V-shaped intervals on the normal data set would be constructed (Altman, 2009; Ludbrook, 2010). To aid interpretation (as there were no normally distributed data for the LOA) BA analysis were also presented with a percentage by means plot. Validity data was analysed only by comparing the most experienced user (tester 2) and the IKD.

5.5 Results

The recruited physiotherapists (4) had between 23 and 13 years of experience. Results are described for 24 participants, 13 females (54%) and 11 males (46%) with the following mean: age of 23.5 (± 5.4); height of 169.39 cm (± 10.72); the weight of 73.74 kg (± 11.83); and BMI of 25.77 (± 4.03).

Twenty-three participants were right hand dominant (96%), and only one participant was left hand dominant (4%). Data was also collected for the size of the moving segment to allow comparison with the IKD. Five participants abandoned the trial before its completion; information on this issue is provided by tester and device (Table 5.3).

Table 5.3 Participants tested by each tester

	T1	T2	T3	T4	IKD
D1	23	23	22	21	19
D2	21	21	NA	NA	NA

Note: T – tester; IKD – Isokinetic Dynamometer; D1 – Day 1; D2 – Day 2; NA – Not applicable

5.5.1 Intra tester reliability and responsiveness

5.5.1.1 *Peak force*

As can be seen in Table 5.4, for tester 1, Intrarater reliability using ICC (2,3) comparing day 1 vs day 2 shows excellent reliability for mean peak force in the shoulder (0.93), elbow flexion (0.93) and knee extension (0.83) and moderate reliability for knee flexion (0.74). Analysis of data from tester 1 revealed a SEM of 0.59kg (8.62%) for shoulder abduction, 0.68kg (7.37%) for elbow flexion, 1.94kg (7.73%) for knee extension and 1.72kg (9.65%) for knee flexion. Whereas, when looking at the MDC values it can be seen that these ranged from 1.63kg (23.89%) on shoulder abduction to 1.90kg (20.40%) on elbow flexion, 5.37kg (21.40%) knee extension and 4.75kg (26.73%) on knee flexion.

Data from tester 2 (table 5.5) revealed an excellent reliability for mean peak force in shoulder (0.93), elbow flexion (0.96) and good reliability for knee extension (0.78) and for knee flexion (0.76). SEM values were: 0.53kg (7.97%) for shoulder abduction, 0.57kg (5.32%) for elbow flexion, 2.31 (8.92%) for knee extension and 1.73 kg (9.21%) for knee flexion. MDC values range from 1.46kg (22.09%) on shoulder abduction, to 1.58 kg (14.73%) on elbow flexion, 6.40 kg (24.72%) knee extension and 4.79kg (25.50%) on knee flexion.

Table 5.4 Mean Peak force and Intra-rater reliability – (p < 0.01) Tester 1

MOVEMENT	D1 (KG (SD))	D2 (KG (SD))	ICC(2,3) (95%CI)	SEM (%)	MDC(%)
SHOULDER ABDUCTION	6.99(2.32)	6.66(2.13)	0.93(0.83-0.97)	0.59(8.62%)	1.63(23.89%)
ELBOW FLEXION	9.19(2.29)	9.38(2.84)	0.93(0.84-0.97)	0.68(7.37%)	1.90(20.40%)
KNEE EXTENSION	24.56(4.09)	25.59(5.24)	0.83(0.62-0.93)	1.94(7.73%)	5.37(21.40%)
KNEE FLEXION	18.06(3.70)	17.49(2.99)	0.74(0.41-0.89)	1.72(9.65%)	4.75(26.73%)

Note: D1 – Day 1; D2 – Day 2; KG- Kilogram; SD – standard deviation, Abd. – Abduction;

Table 5.5 Mean Peak force and Intra-rater reliability – (p < 0.01) Tester 2

MOVEMENT	D1 (KG (SD))	D2 (KG (SD))	ICC (95%CI)	SEM (%)	MDC(%)
SHOULDER ABDUCTION	7.09(2.08)	6.17(1.91)	0.93(0.43-0.98)	0.53(7.97%)	1.46(22.09%)
ELBOW FLEXION	10.98(3.01)	11.01(2.66)	0.96(0.89-0.98)	0.57(5.32%)	1.58(14.73%)
KNEE EXTENSION	26.71(4.89)	25.06(4.96)	0.78(0.49-0.90)	2.31(8.92%)	6.40(24.72%)
KNEE FLEXION	17.83(2.96)	19.77(3.53)	0.76(0.20-0.91)	1.73(9.21%)	4.79(25.50%)

5.5.1.2   *Angle of peak force*

The Mean angle of peak force has moderate reliability for elbow flexion and knee flexion for tester 1 (0.61 and 0.62 respectively – see table 5.6) - whereas shoulder abduction and knee flexion have poor reliability. For tester 2 (table 5.7) knee extension and knee flexion revealed moderate reliability (0.74), whereas shoulder abduction and elbow flexion had poor reliability (ICC <0.5).

Table 5.6 Mean Angle of Peak Force and Intrarater reliability – (p < 0.01) Tester 1

MOVEMENT	D1 T1 (º)	D2 T1 (º)	ICC (95%CI)	SEM (% SEM)	MDC (% MDC)
SHOULDER ABD.	24.16(10.62)	30.36(13.56)	0.17(-0.77-0.63)	11.07(40.62%)	30.68(112.53%)
ELBOW FLEXION	82.49(20.75)	82.21(12.70)	0.61(0.114-0.83)	10.70(12.99%)	29.64(35.99%)
KNEE EXTENSION	58.59(20.44)	61.61(13.60)	0.25(-0.71-0.68)	11.12(18.50%)	30.80(51.24%)
KNEE FLEXION	30.64(22.47)	35.13(19.72)	0.62(0.129-0.83)	13.03(39.63%)	36.10(109.78)

Notes: Abd.- Abduction; D1 – Day 1; D2 – Day 2; T1 -Tester 1; T2- Tester 2

Table 5.7 Mean Angle of Peak Force and Intrarater reliability – (p < 0.01) Tester 2

MOVEMENT	D1 T2 (º)	D2 T2 (º)	ICC (95%CI)	SEM (% SEM)	MDC(%MDC)
SHOULDER ABD.	14.98(5.51)	25.31(11.35)	0.27(-0.28-0.64)	7.60(37.74%)	21.06(104.53%)
ELBOW FLEXION	91.14(25.92)	97.31(16.20)	0.21(-0.81-0.66)	19.21(20.39%)	53.22(56.48%)
KNEE EXTENSION	54.21(24.22)	67.26(25.26)	0.74(0.41-0.89)	12.88(21.21%)	35.68(58.74%)
KNEE FLEXION	37.43(21.87)	59.38(19.60)	0.45(-0.18-0.76)	15.40(31.81%)	42.65(88.11%)

Notes: Abd. – Abduction; D1 – Day 1; D2 – Day 2; T1 -Tester 1; T2- Tester 2

5.5.1.3　*Angular impulse*

Angular impulse data reliability for the comparison between-sessions for tester 1 and tester 2 is presented in table 5.8 and 5.9. Results for angular impulse were good for elbow flexion but poor for all other movements. Elbow flexion showed a SEM% of 10.65 and MDC% of 29.50. Whereas results were superior for tester 2 with good reliability for shoulder abduction and elbow flexion. Moderate for knee flexion and poor for knee extension. All MDC values were consistently higher than 40% for all movements with ICC>0.50.

Table 5.8 Angular Impulse Intratester reliability between day– Tester 1

MOVEMENT	ICC (95% CI)	SEM (%)	MDC (%)
SHOULDER ABDUCTION	0.25 (-0.26 – 0.59)	NA	NA
ELBOW FLEXION	0.79 (-0.05-0.94)	3.79(10.65)	10.50(29.50)
KNEE EXTENSION	0.27 (-0.17 – 0.59)	NA	NA
KNEE FLEXION	0.46 (-0.03 – 0.73)	NA	NA

Table 5.9 Angular Impulse Intratester reliability between day– Tester 2

MOVEMENT	ICC (95% CI)	SEM (%)	MDC (%)
SHOULDER ABDUCTION	0.84 (0.61-0.94)	4.49 (17.02)	12.45 (47.14)
ELBOW FLEXION	0.83 (0.58-0.93)	5.47 (15.96)	15.14 (44.20)
KNEE EXTENSION	0.39 (-0.51-0.75)	21.70 (23.31)	60.11 (64.56)
KNEE FLEXION	0.69 (0.25-0.88)	11.15 (15.81)	30.88 (43.79)

5.5.2   Inter-rater reliability and responsiveness

As shown in table 5.10, intertester reliability (ICC 2,3 – mean of three repetitions) demonstrated excellent reliability for shoulder and elbow test (0.93 and 0.96), whereas good reliability was depicted for lower limb movements - 0.76 and 0.68 - for knee extension and knee flexion respectively. SEM was lowest for elbow flexion with 0.72kg (6.45%) with an MDC of 2kg (17.87%). These values increased slightly with shoulder abduction 0.60kg (8.78%) and MDC of 1.65kg (24.31%). Knee extension and knee flexion had an SEM of 4.54kg (9.37%) and 2.58kg (13.19%), respectively and an MDC for knee extension of 6.17kg (25.95%) and knee flexion 7.14kg (36.53%).

The angle of peak force for all testers (table 5.11) ranged from poor on knee extension (0.28) and on shoulder abduction (0.43) to moderate in knee flexion (0.55) and good in elbow flexion (0.77). With a minimal SEM of 9.28º (39.93%) for abduction, 9.98º (12.59%) for elbow flexion, 15.14º (30.20%) for knee flexion and 15.84º (35.42%). Due to the poor to moderate reliability MDC values are high, particularly for shoulder abduction 25.70º (110.60%), knee extension 43.86º (98.12%), knee flexion 41.93º (83.64%) and elbow flexion 27.65º (34.89%).

Table 5.10 Mean force per day and Peak force Inter-tester reliability

MOVEMENT	D1 T1 (KG)	D1 T2 (KG)	D1 T3 (KG)	D1 T4 (KG)	ICC (95% CI)	SEM (% SEM)	MDC (% MDC)
SHOULDER ABD	6.99 (2.32)	7.09 (2.08)	6.47 (2.05)	6.62 (2.53)	0.93 (0.86-0.97)	0.60 (8.78%)	1.65 (24.31%)
ELBOW FLEX	9.19 (2.29)	10.98 (3.01)	12.61 (4.29)	12.03 (4.42)	0.96 (0.90-0.98)	0.72 (6.45%)	2.00 (17.87%)
KNEE EXT	24.56 (4.09)	26.71 (4.89)	20.89 (4.18)	22.91 (4.95)	0.76 (0.47-0.89)	4.54 (9.37%)	6.17 (25.95%)
KNEE FLEX	18.06 (3.70)	17.83 (2.96)	20.92 (4.09)	21.42 (6.63)	0.68 (0.42-0.84)	2.58 (13.19%)	7.14 (36.53%)

Table 5.11 Angle of Peak force inter tester reliability

MOVEMENT	D1 T1 (°)	D1 T2 (°)	D1 T3 (°)	D1 T4 (°)	ICC (95% CI)	SEM (% SEM)	MDC (% MDC)
SHOULDER ABD.	24.16 (10.62)	14.98 (5.51)	31.54 (18.18)	22.26 (11.54)	0.43 (0.04-0.71)	9.28 (39.93%)	25.70(110.6%)
ELBOW FLEX.	82.49 (20.75)	91.14 (25.92)	78.83 (15.27)	64.55 (19.93)	0.77 (0.57-0.89)	9.98 (12.59%)	27.65 (34.89%)
KNEE EXT.	58.59 (20.44)	54.21 (24.22)	31.95 (12.04)	34.06 (15.30)	0.28 (-0.05-0.58)	15.84 (35.42%)	43.86 (98.12%)
KNEE FLEX.	30.64 (22.47)	37.43 (21.87)	73.67 (22.67)	58.76 (23.22)	0.55 (0.15-0.78)	15.14 (30.20%)	41.93 (83.64%)

Note: Abd. – Abduction; Flex. – Flexion; Ext. – Extension; D1- Day 1; T1 – Tester 1; T2 – Tester 2; T3 - Tester 3; T4 – Tester 4

Intertester reliability for angular impulse (table 5.12) was excellent for elbow flexion but lower (good) for shoulder abduction and moderate for knee extension and flexion. In table 5.12 SEM in percentage is lowest (11.60%) for elbow flexion, followed by knee extension (12.51%), knee flexion (14.49%) and shoulder abduction (19.50%). MDC, which relates to the minimal detectable error one can expect to measure without error, ranged from 32.13% for elbow flexion, 34.65% for knee extension, 40.13% for knee flexion and shoulder abduction for 54.02%.

Table 5.12 Inter-tester reliability angular impulse -All testers

MOVEMENT	ICC (95%CI)	SEM (% SEM)	MDC (% MDC)
SHOULDER ABD	0.82 (0.64-0.93)	6.73 (19.50)	18.65 (54.02)
ELBOW FLEX	0.92 (0.84-0.97)	5.72 (11.60)	15.86 (32.13)
KNEE EXT	0.74 (0.46-0.89)	12.91 (12.51)	35.75 (34.65)
KNEE FLEX	0.74 (0.48-0.89)	12.66 (14.49)	35.07 (40.13)

5.5.3   Validity

Validity data is presented only for the comparison with the most experienced user with the dynamometer – tester 2. The results from IKD vs Tester 2 (table 5.13) for the peak torque analysis show strong correlations for all movements, with higher results (0.88) for shoulder abduction and elbow flexion. However, poor correlations were reached for the angle of peak torque in all tested movements (table 5.14). WLP regression for angle of peak torque was not performed as the p-value for the correlation are superior to 0.05 (table 5.14) for every comparison, which means there is no acceptable level of correlation between the two variables, and therefore they are not suitable for WLP regression (Ludbrook, 2012).

Table 5.13 Correlation IKD vs ASSA – Peak torque (Tester 2)

	PEARSON'S R	P-VALUE	95% CI	T-VALUE
SHOULDER	0.88	<0.01	0.70-0.95	7.45
ELBOW	0.88	<0.01	0.71-0.95	7.65
KNEE EXT.	0.72	<0.01	0.39-0.88	4.23
KNEE FLEX.	0.66	<0.01	0.29-0.86	3.60

Table 5.14 Correlation IKD vs ASSA – Angle of peak torque (Tester 2)

	PEARSON'S R	P-VALUE	95% CI	T-VALUE
SHOULDER	-0.05	0.85	-0.49-0.42	-0.19
ELBOW	0.12	0.64	-0.36-0.54	0.48
KNEE EXT.	-0.33	0.17	-0.68-0.15	-1.42
KNEE FLEX.	-0.18	0.47	-0.58-0.30	-0.74

Table 5.15 WLP regression IKD vs Tester 2

MOVEMENT	ELEVATION (95% CI)	SLOPE (95% CI)
SHOULDER ABDUCTION	13.42 (8.14-17.55)	1.32 (1.01-1.69)
ELBOW FLEXION	7.00 (1.49-11.36)	0.74 (0.58-0.94)
KNEE EXTENSION	36.63 (9.91-55.46)	0.44 (0.31-0.63)
KNEE FLEXION	26.69 (8.66 – 39.06)	0.51 (0.35-0.75)

WLP regression for peak torque (table 5.15) shows both fixed and proportional bias in all movements. In the following sections (section 5.5.3.1 to 5.5.3.4), LOA results for several movements are presented.

5.5.3.1 *Shoulder Abduction LOA*

Shoulder abduction was overestimated on average by 88.48% (LOA: 16.15%-160.80%) (table 5.16). Figure 5.8 shows that means lower than 25Nm there is an overestimation

(from the prototype) of around 125% and higher means (above 25 Nm), overestimating the gold standard by around 50% (figure 5.8).

Figure 5.7 WLP regression peak torque shoulder abduction
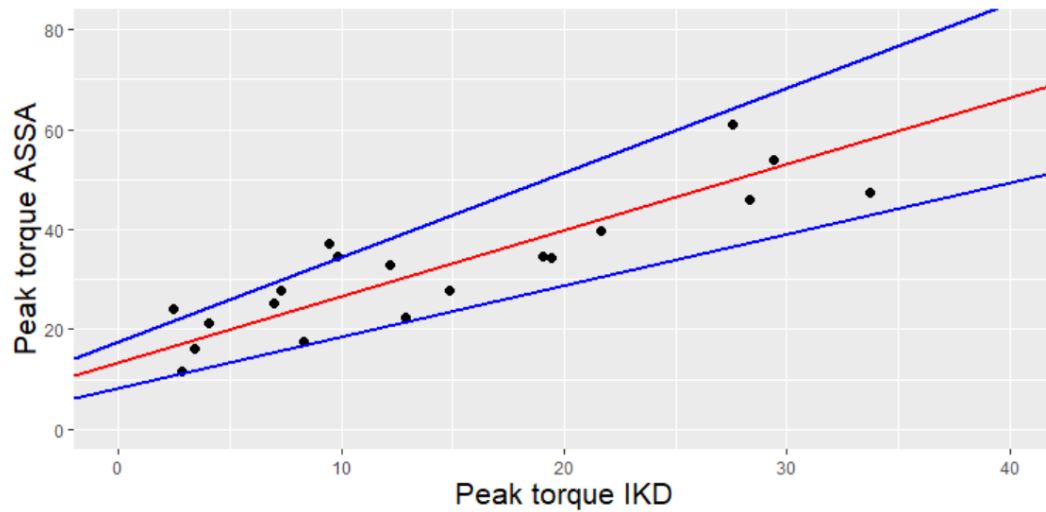
Figure 5.8 Bland-Altman plot shoulder abduction in Nm
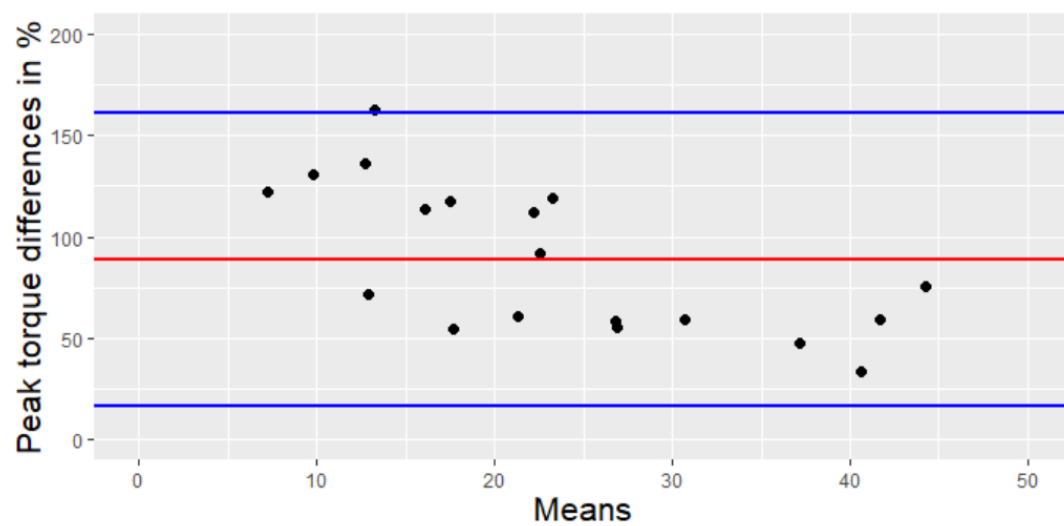
Table 5.16 Bland-Altman LOA Shoulder abduction -ASSA vs IKD in Nm (percentage)

	ESTIMATE (%)	LOWER CI (%)	UPPER CI (%)
BIAS	88.48	71.43	105.52
LOWER LOA	16.16	-13.36	45.69
UPPER LOA	160.80	131.27	190.32

5.5.3.2 *Elbow flexion LOA*

Elbow flexion was overestimated on average by 2.85% (LOA: -35.32% - 38.17%), with a tendency for lower means to be overestimated and higher means to underestimate the IKD (table 5.17 and figure 5.10). WLP regression is presented in figure 5.9.

Table 5.17 Bland-Altman LOA Elbow flexion -ASSA vs IKD in Nm (percentage)

	ESTIMATE (%)	LOWER CI (%)	UPPER CI (%)
BIAS	2.85	-5.70	11.40
LOWER LOA	-35.32	-18.62	-48.26
UPPER LOA	38.17	24.33	53.96

Figure 5.9 WLP regression elbow flexion in Nm
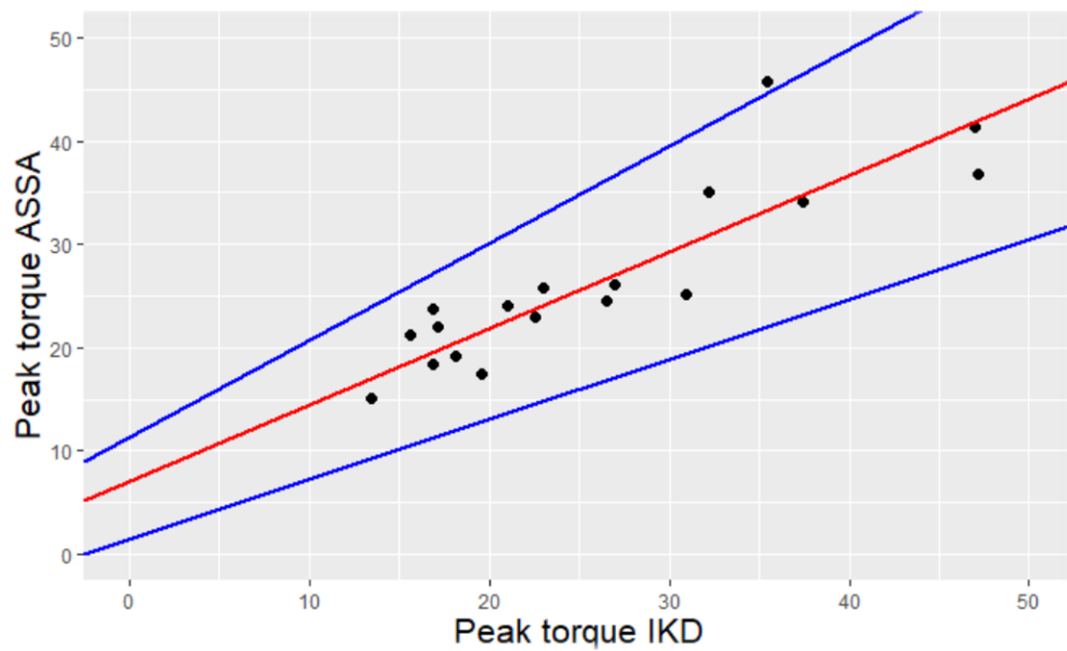
Figure 5.10 Bland-Altman plot Elbow flexion (percentage by means in Nm)
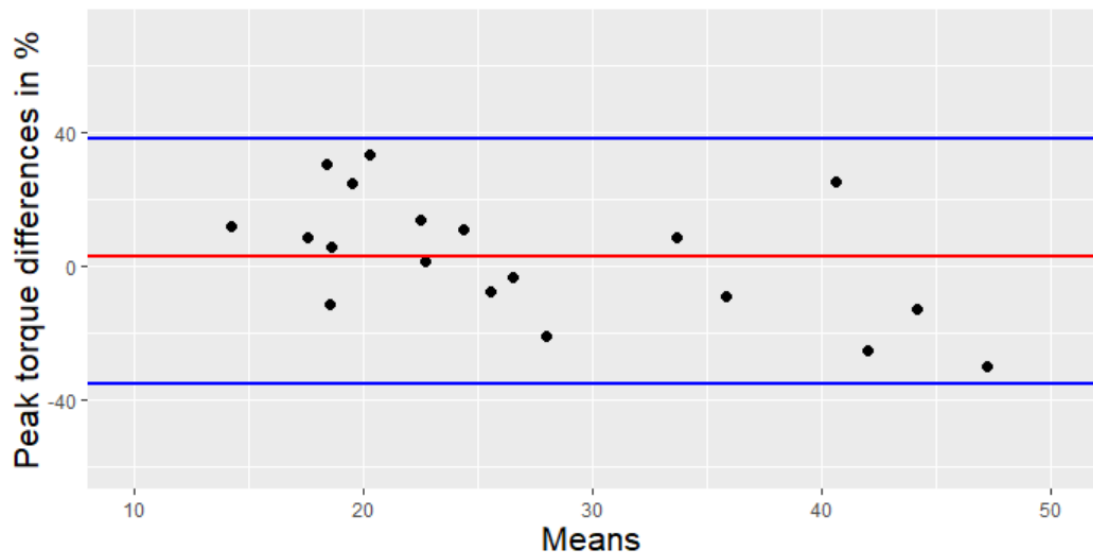
5.5.3.3   *Knee Extension*

Knee extension was underestimated on average by -33.10% (LOA: -70.83% - 4.62%), with most data points showing it underestimates the criterion measure from -5% to -50% (table 5.18 and figure 5.12). WLP regression for knee extension is presented in figure 5.11.

Table 5.18 Bland-Altman LOA knee extension – ASSA vs IKD (percentage)

	ESTIMATE (%)	LOWER CI (%)	UPPER CI (%)
BIAS	-33.10	-42.00	-24.21
LOWER LOA	-70.83	-86.23	-55.43
UPPER LOA	4.62	-10.78	20.02

Figure 5.11 Knee extension WLP regression in Nm
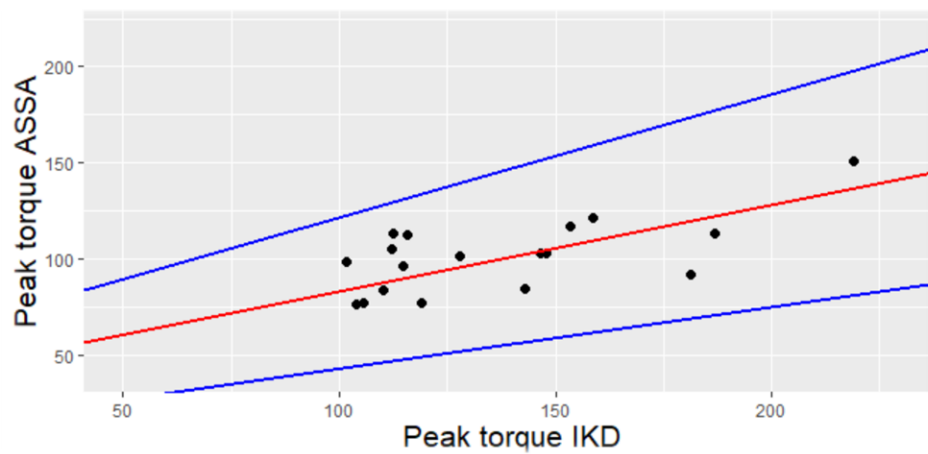
Figure 5.12 Bland-Altman plot Knee extension (percentage by means in Nm)
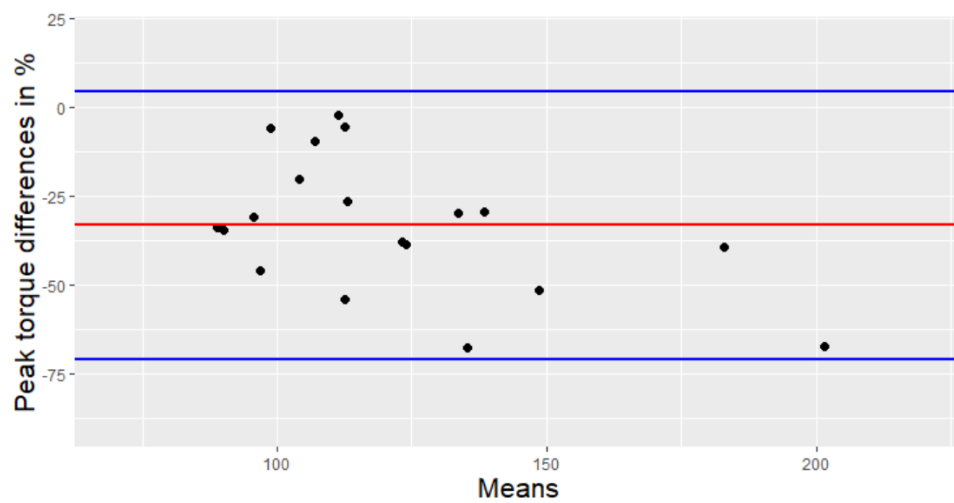
5.5.3.4 Knee Flexion

ASSA underestimated knee flexion on average by -12.85% (LOA: -52.11% - 26.42%) (table 5.19 and figure 5.14), with higher mean values of torque closer to the lower LOA, the WLP regression is presented in figure 5.13.

Table 5.19 Bland-Altman LOA Knee flexion – ASSA vs IKD (percentage)

	ESTIMATE (%)	LOWER CI (%)	UPPER CI (%)
BIAS	-12.85	-22.10	-3.59
LOWER LOA	-52.11	-68.14	-36.08
UPPER LOA	26.42	10.39	42.45

Figure 5.13 WLP regression knee flexion in Nm
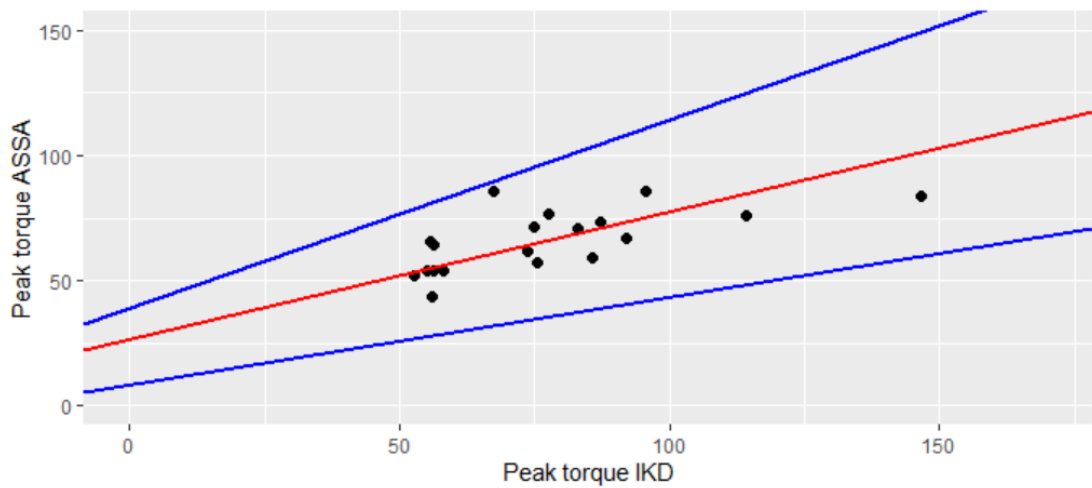
Figure 5.14 Bland-Altman plot Knee flexion (percentage by means in Nm)
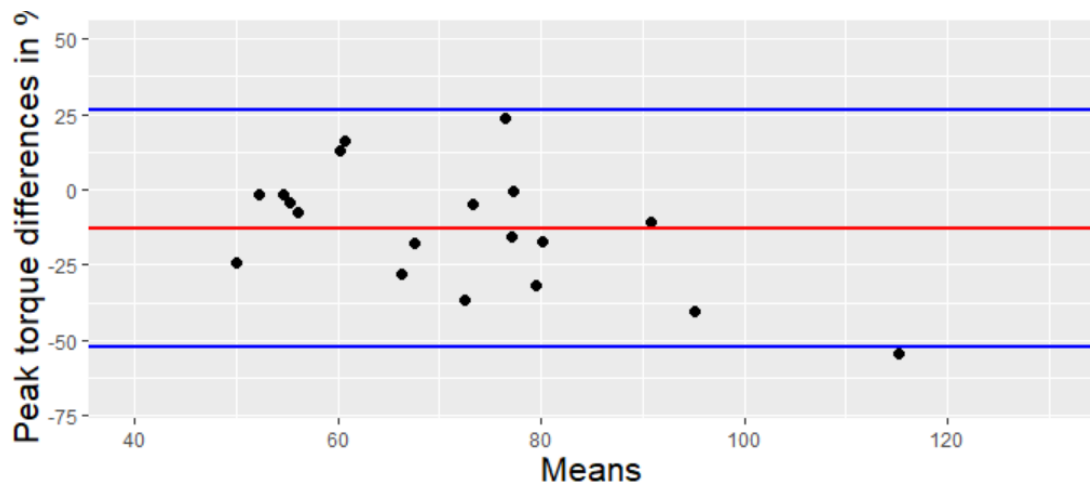
5.6 Discussion

The use of concentric tests when using an HHD is not usual in physiotherapy practice, but the current decrease in costs warrants the investigation for the development of a device that most physiotherapists can use. Results demonstrated good to excellent for both intra and intertester reliability in peak force detection. At the same time, angle of peak force had moderate to good reliability in the elbow and knee flexion for tester 1 and moderate for tester 2 on knee extension. Validity analysis shows high correlation between ASSA and the IKD but there is a fixed and proportional bias when using the device, particularly for larger forces in each movement. Due to the absence of literature regarding the use of concentric movements for manual muscle tests, comparisons with isometric HHD are limited.

5.6.1 Intra tester between-session reliability

5.6.1.1 Peak force

In terms of peak force reliability (between-session), results show an excellent correlation for shoulder abduction in both testers. Which is similar to a comparable device and concentric movement investigated by Cadogan, et al. (2011). The current results are also superior to the reliability reported by Dollings, et al. (2012), who tested intra and intertester reliability of the PowerCommander II in clinical tests performed on the shoulder. The SEM was 5.9N (0.59kg) (tester 1) and 5.3N (0.53kg) (tester 2), where theirs

was between 8.45-9.34N (both limbs tested). However, they used an isometric test and their resistance point was above the elbow joint (as did Cadogan, et al., 2011), which might make it more difficult for the therapist to resist due to a smaller lever - which could increase SEM.

Vermeulen, et al. (2005) when investigating the use of a fixed HHD reported elbow flexion with an ICC of 0.95 and 0.96 for HHD and fixed HHD, respectively, while shoulder abduction was 0.86 and 0.82, respectively. These values are similar to the results presented in this thesis for elbow flexion but lower than the concentric approach (0.93 – tester 1; 0.96 – tester 2) for shoulder abduction. The same authors also state some discomfort when using the device, which was not reported with this prototype.

Knee extension reliability revealed good results for both testers with SEM ~7-9%. When compared with research performed by Martins, et al. (2017), whom used a fixed HHD, their reliability was excellent, yet, their SEM was larger (12% right side; 20% left side), where the concentric movement was 0.83 for tester 1 (SEM=7.73% and MDC=21.4%) and 0.78 for tester 2 (SEM=8.92% and MDC=24.72%). The device and the concentric approach appears to have lower reliability for a group of healthy subjects but better SEM, however, it is unknown how these values would differ in specific conditions.

Reliability regarding knee flexion from Martins, et al. (2017) were moderate (ICC: 0.66-0.62) with and SEM of 20-22% whereas these findings are slightly better with an ICC of 0.74 for tester 1 (SEM=9.65%; MDC=26.7%) and for tester 2 ICC was 0.76 (SEM=9.21%; MDC=25.50%). Their testing position uses the chair and straps from an IKD which might influence their results by providing stability to the participant - that was inexistent in this trial (see section 5.3.4). The knee flexion was tested with the participant in a sitting position (same as knee extension), and the participant was asked to flex their knee. This was chosen to compare data with the IKD. However, this position would not normally be selected in a clinical setting as it allows for neither good segment control nor patient stability, which might be why the ICC was lower than expected.

5.6.1.2 *Angle of peak force*

Angle of peak force was poor in shoulder abduction for between-session reliability (both testers). This might be due to the large range of motion and the testing position. The position was selected to be as similar as possible with the testing position used in the IKD (with the resistance applied on the subject's hand), this might have impaired the ability of the therapists to resist the movement throughout the range and a different position should be tested in further studies. The preferred position was suggested by Bohannon (1997) where the resistance should be applied above the elbow – this should be investigated in future research.

Similar problems might have a causal relationship for the fact that none of the movements had an ICC higher than 0.75 for both testers. Further exploration in terms of therapist position and range of motion for each joint should be considered in the future. Besides this, SEM and MDC values should be carefully considered as they might still be relevant to practice if the reliability is deemed improvable or if the clinical situation is deemed appropriate. For instance, when one expects a considerable improvement of ROM, it might still be acceptable to have a tool with a large MDC.

For the first time, it was documented that the angle of peak force can have moderate reliability (elbow flexion and knee flexion for tester 1 and knee extension for tester 2) when using the concentric approach. In the next chapter this thesis will investigate if the use of sonification can improve this outcome.

5.6.1.3 *Angular impulse*

Angular impulse data retrieved from concentric manual tests have never been published to the author's knowledge. The results present in this thesis demonstrate that the concentric approach displays moderate to good reliability for tester 2 (the exception being knee extension between-session 1 and 2). However, results for between-session in regards to tester 1 are poor with the exception of elbow flexion. Tester 1 has extensively used MMT and HHD, but results are not enough to ascertain that angular impulse can provide a dynamic assessment of strength for between-session reliability. Considering that there is no previous data from HHD regarding this outcome, it might be argued that this has limited practical ability, and the MDC's are too large (over 30% for all tests). Nonetheless,

this approach should be explored in different joints and populations to investigate further how physiotherapy and rehabilitation can benefit this parameter if deemed reliable.

5.6.2 Inter tester reliability

5.6.2.1 *Peak force*

Peak force results for all testers demonstrated excellent reliability for shoulder and elbow movements, good reliability for knee extension, and moderate knee flexion reliability. For inter-tester reliability results demonstrated that the concentric approach appears to be superior in a young and healthy population. Results from Clarke, et al. (2011) displayed moderate inter-tester reliability (0.61 – day 1 and 0.66 – day 2) with a slightly younger population (21.8 ± 2.4 years) than the one presented in this section (23.5 ± 5.4 years). While shoulder abduction from the Dollings, et al. (2012) research presented moderate reliability (0.77 – right side; 0.87 – left side) with an SEM of 15.1N for the right side and 13.8N for the left side, the concentric method displayed a higher ICC (0.93), while demonstrating a lower SEM (0.60Kg – 5.89N). These differences might arise from the testing position and dynamometer placement, which in Dollings, et al. (2012) is on the humerus – which means it might be more difficult to resist than the movement in this chapter due to a longer lever(resistance applied distally on the upper limb). These same authors also reported elbow flexion data with excellent reliability for both sides tested (0.91-0.93), slightly lower than the displayed by ASSA. Nonetheless, they have a wider 95%CI (0.68-0.97; 0.83-0.97) than the concentric test (0.90-0.98). Reported SEM was close to 20N for both sides (20.5N – 19.6N), where the SEM in the current research was 7.06N.

There were also encouraging results from the concentric approach as the inter-rater reliability for knee extension was good (>0.75). At the same time, Clarke, et al. (2011) reported moderate inter-tester reliability (0.61-0.66) in two different days with a population similar to the present in this chapter. These are particularly encouraging results from one of the most difficult joints to assess due to the large torque values.

Thorborg, Bandholm and Hölmich (2013) while investigating the use of a fixed HHD, reported better reliability values than the prototype using the concentric approach. Their

ICC (2,1) was 0.84 for knee flexion, with an SEM of 27.1N (SEM 9%) and MDC 75.1N (24.8%), where ASSA displayed 0.68 ICC with a 13.19% SEM and 36.53% MDC. They did use two physiotherapy students, but this should not significantly impact their results compared to ASSA as they did not apply the resistance. As referred in the previous section (5.1.1), this might be due to the lack of stabilisation in the selected position and should be assessed further.

5.6.2.2 *Angle of peak force*

The angle of peak force displayed poor intertester reliability for both shoulder abduction and knee extension, moderate for knee flexion and good for elbow flexion reliability. The fact that two hands are used in the elbow flexion test might facilitate the appropriate technique, but the other joints do not display acceptable reliability. This means that ASSA cannot consistently detect peak angle in most joints tested – with different testers - even when using experienced testers in a young and healthy population. Nonetheless, it exhibits an inappropriate amount of error for a clinical setting in most joints. If the improvements in software and technique are enhanced in the future, then the device might be used to monitor and aid in the progress of treatment.

5.6.2.3 *Angular impulse*

Angular impulse results for inter-tester reliability reveal a promising feature of the new prototype with moderate to excellent degrees of reliability for the tested joints while SEM was below 20% for all movements tested. Angular impulse data has not been published anywhere else to the author's knowledge but can be used as a measure of endurance. This has been explored in the literature when using IKDs, but usually, sports scientists tend to prefer peak force or work data as ways of analysing muscle performance. For physiotherapists, however, it might be a relevant measure of force in a clinical setting.

5.6.3 Validity

5.6.3.1 *Peak torque*

The results demonstrate that the device is valid. There was a moderate to strong correlation on shoulder abduction; elbow flexion; knee extension and knee flexion

between ASSA and the IKD. Lower correlation from knee flexion might have arisen from weak stabilisation and segment control, and therefore, this position is not recommended for knee flexion testing, and an alternative procedure should be found for the concentric test.

WLP regression analysis demonstrates fixed and proportional bias in all movements tested (table 5.15), which means the prototype cannot replace the IKD, but the current data can be used to calibrate the device, improving the agreement between methods.

In their 1999 paper, Bland and Altman stated that "agreement is not something which is present or absent but something which must be quantified" (Bland and Altman, 1999, p.159). In this chapter, results displayed acceptable LOA for elbow flexion, knee extension and knee flexion. LOA were too wide to allow for the substitution of one device by the other (which is many times the goal of BA analysis) but enough to warrant that the prototype is measuring the same construct as the gold standard. Testing from the IKD was reported by the participants in the shoulder abduction testing as problematic and they reported difficulties in performing to the best of their ability, which might also affect the results for the validity analysis. This issue is relevant for both IKD data and HHD as the testing position does not allow for great mechanical advantage at the end of the range of movement. Despite this, the results gathered from validity analysis between the HHD and IKD can be used to improve the data output of the newly developed prototype. For instance, using a calibration formula obtained from the data collected to increase agreement between ASSA and IKD.

5.6.3.2   *Angle of peak torque*

Data from the angle of peak torque show a poor correlation between ASSA and the IKD. This means the concentric technique tested here does not allow for the comparison of the angle of peak torque, and the device should not be used for that purpose. In previous research by Cadogan, et al. (2011), when testing an HHD with the ability to detect ROM, their results show good reliability in active shoulder abduction movements, but they did not assess the device's ability to detect angle peak torque. In future endeavours, it might be necessary to use sensors independent from the HHD's position as in the device developed by Li, et al. (2006) to evaluate their performance for the angle of peak torque.

154

5.6.3.3 *Angular Impulse*

The validity of angular impulse was not investigated in this thesis. This was due to the fact that angular impulse data originates from the force applied for a certain period of time (in Nm*sec), however, the method used did not account for differences in ROM for each of the tested movements and, considering the muscle length-tension relationship, this would make it inappropriate to compare force production across the available range of motion. One possible solution for future research would be to attach the prototype to the IKD while performing the tests, this could provide valuable information regarding the not only angular impulse but also angle of peak torque.

5.6.4 Strengths and limitations

The results presented in this section show for the first time that a concentric approach to manual muscle testing using an HHD is a reliable source of peak force in all movement tested for a group of experienced testers. Tough, knee flexion and knee extension movements just below the cut-off values of 0.75 for reliability but with narrow CI to assume that if increased stability was provided as explained previously the reliability would be above 0.75. Whereas angular impulse demonstrated good reliability for inter-tester reliability for upper limb movements and knee extension, but below acceptable for lower knee flexion. As stated before, a different testing position should increase this. ASSA v3 is also able to provide valid output regarding peak force.

It is yet unknown how less experienced testers and female testers will perform using this approach. But given that peak force in concentric testing is decreased when compared to isometric testing, this might facilitate manual muscle tests for those specific groups of physiotherapists. This will be investigated in the following chapter.

Data collected shows that HHD does not show agreement to the point that allows for replacing the gold standard approach. The author argues that this should not be expected, considering the resources involved in each of them. Nevertheless, a therapist should decide if the reported LOA are acceptable for their line of work, i.e., target population, available resources and country where they practice.

Lastly, this version of the device needs to undergo further software reliability testing with VC as the users reported a failure to initiate the procedure two to four times every day

before testing commenced where the device had to be restarted – this means data collection was not affected but it results in loss of time for researcher and participants.

5.7 Conclusion

The new prototype is reliable in conveying peak force data for the same tester (as also shown in chapter 4) and for different testers in elbow flexion, shoulder abduction, knee extension and knee flexion. Angular impulse displayed mixed results in between session-intra-tester reliability for shoulder abduction and knee movements but moderate to good inter-tester reliability.

The device demonstrates a good degree of correlation and agreement with the gold-standard to detect peak torque but not enough to replace the gold-standard. It can still be considered valid. On the other hand, regarding the angle of peak force, the correlation data show it is not a valid instrument and does not seem to convey reliable data to be used in clinical practice for that purpose.

6. SONIFICATION TESTING AND RELIABILITY COMPARISON BETWEEN PHYSIOTHERAPISTS WITH MORE THAN 15 YEARS OF EXPERIENCE AND PHYSIOTHERAPY STUDENTS

6.1 Introduction

Evidence-based clinical approaches require constant reasoning and improvement of daily practices. Developing reliable measurement tools that suit both experienced and inexperienced physiotherapists is intertwined with improving physiotherapy interventions. The new prototype was tested in the previous chapters with a homogenous group of experienced testers and quantified its validity, reliability and responsiveness, a further step into the device development was needed. Physiotherapists with similar levels of experience showed good reliability and responsiveness, however it has not been investigated how different characteristics such as experience might affect reliability and responsiveness. Physiotherapists who use dynamometers have different experience levels and strength. Therefore, this chapter investigates how reliability and responsiveness is affected in a heterogeneous group of physiotherapists with distinct genders and experience levels. The difference in tester strength, which might be related to biological sex, changes muscle test output results using regular HHDs (Clarke, et al., 2011). At the same time, experience does not appear to influence the results in isometric testing (Bohannon and Wilkhom, 1992; Goonetilleke, et al., 1994; Grooten and Äng, 2010; Keep, et al., 2016) and appears to be less relevant than tester strength; however, this needs to be examined when using a concentric approach to manual muscle tests. It is also unknown how the sonification of manual muscle tests affects such tests' reliability and force production, which will be explored here.

6.2 Objective

This chapter investigates how sonification, testers' experience, and biological sex influence reliability in upper and lower limb manual muscle tests. The current chapter's

objective is to quantify the reliability of ASSA v4 with and without sonification using physiotherapists and physiotherapy students as testers. This was done by comparing elbow flexion and knee flexion concentric manual muscle tests with and without sonification as audio-feedback.

6.3 Methods

6.3.1 Prototype evolution – Version 4

As previously explained in chapter 3, the improvements regarding version 4 related to the software and the amount of information displayed to the physiotherapist. Namely, data regarding the instant acceleration (for segment movement) was made available, and sonification provided live audio-feedback to the tester through earphones. The hardware did not suffer any changes from the previous data collection, while sonification was added to the software.

6.3.2 Design

A repeated-measures single-blinded randomised controlled trial was planned with four testers and 24 participants. Each tester would assess both elbow flexion and knee flexion in all participants three times under different conditions (control - no sound; intervention 1 - sound 1; intervention 2 - sound 2).

Training in the use of ASSA was provided with and without sonification. A similar and consistent verbal incentive was taught and provided by all physiotherapists and physiotherapy students during practice and testing. The training was given for two hours (with at least 20 minutes for MMT with sonification), but no proficiency test was arranged. This was a deliberate choice as in clinical practice, physiotherapists will not have that option. By doing so, the author hypothesises that the results will be similar to the ones found in a clinical setting across different sites.

All testers were able to trial the HHD with sonification while using earphones (VKUSRA® - wireless headphones), and each mapping was explained to the users. Volume was selected by each of the participants for maximum comfort so that all sound

range was audible. No sound was perceptible to the participant on whom the test was being performed, isolating the sound-feedback to the tester only. During testing, as in previous chapters, the physiotherapists were blinded to the force/angle/speed results to prevent bias – however, they could not be blinded to the sound/no sound condition. Conditions were randomised to reduce learning and carry-over effects. As the testers worked in pairs, the first tester to begin the MMT was also randomised.

Due to a technical problem with the software update for this chapter, sonification testing had to be stopped on day 3 of data collection. The problem originated on the code that attributed different notes to each of the variables (force or speed) and on every attempt to run the program, it would stop, and testers would be required to restart the process. When the problem arose, only nine participants had completed the data collection for sound condition 1. At the time, due to family commitments, the author was in the UK and although several attempts were made, the software error could not be solved online. This meant that further testing was altered, and testers were instructed to proceed with manual muscle testing without sound feedback until the author was back in Portugal to provide technical assistance. In sum, 23 participants were tested without sound, and only nine were tested with one of the sound conditions (sound 1). Unfortunately, once the author was back in Portugal and the sonification problem was solved, COVID-19 was starting to impact Portugal, and this research was halted. Nonetheless, *post-hoc* analysis demonstrated that the results for the sonification were statistically significant, and large effect sizes were found, which meant meaningful information was collected from the sample of nine participants.

6.3.3 Recruitment and participants

Physiotherapists and physiotherapy students were recruited following judgement sampling from a pool of available participants from HSHPRC: two experienced physiotherapists with more than ten years of experience in manual muscle tests were recruited to serve as experienced testers. In contrast, two physiotherapy students (third year) were recruited to be novice testers. In order to guarantee a similar approach in each

movement, correct procedural positions were also taught to all testers before the data collection commencement.

Criteria for inclusion and exclusion were defined for both testers (experienced physiotherapists and physiotherapy students) as well as participants. The sample size for reliability was determined considering a power of 90% and alpha of 0.05, for a difference of acceptable reliability (0.70) to expectable reliability (0.90) this research would need to recruit at least 18 individuals with three observations per participant (Bujang and Baharum, 2017). As done previously and since some data may be lost or participants may withdraw, 20%-30% more participants were recruited - a total of 24 subjects was then included.

Criteria were established for recruitment for both physiotherapists and participants. Different inclusion criteria were created for each group of testers: 1) Experienced Physiotherapists - More than ten years clinical experience; Ability to perform a manual muscle test against strong resistance; 2) Physiotherapy undergraduates - Less than 4 years of experience of manual testing experience; Ability to perform a manual muscle test against strong resistance; Over 18 years of age. Exclusion criteria for both groups of physiotherapists: Hearing deficit; Inability to understand the protocol; Inability to understand English written/spoken. In the end, two experienced (one male and one female) physiotherapists and two physiotherapy undergraduates (one male and one female) were recruited.

Separate criteria were created for the participants. Inclusion Criteria: Age between 18 and 60 years; No major surgery/injury in the last six months; No current injury on back or limbs. Exclusion criteria for participants: Recent injury or disability that prevents the subject from performing a maximum voluntary contraction (upper and lower limbs); Inability to provide consent or understand the procedure.

Recruited physiotherapists had different levels of experience. The most experienced between 16 and 18 years of experience and the physiotherapy students were both third-year students.

In order to minimise fatigue, participants had a minimum of 48h between testing and were tested at similar times of the day. Ethics approval was required in Portugal, and the United

Kingdom for data collection. All participants were given a consent form to sign before data collection – ARU Ethics Code ESPGR-08 (Appendix 9) as well as HSHPRC, where the data was collected.

6.3.4 Testing procedure

In line with previous data collection, similar methods were established (see chapter 5 section 3.4). According to personal preference, a warm-up with 1-3 repetitions was permitted (at 25-50% maximum voluntary contraction). Participants were given five minutes rest between testers to avoid fatigue (Martins, et al., 2017). On table 6.1 testing positions are presented.

As in the previous testing with the HHD, participants were asked to refrain from vigorous physical activity 48h before each testing session and maintain their nutritional and activity habits as described in previous chapters to minimize fatigue and changes in strength condition. Subjects were, as much as possible, scheduled to perform the test at similar times of the day to minimise any diurnal influence in strength production. The participants' dominant side was determined by asking the participant which leg they used to kick a ball. Only the dominant side was tested.

6.3.5 ASSA protocol

Every tester used the same mark on the participants' tibia or forearm to position the centre of the dynamometer every single session – marking was recreated every testing day using individual measures recorded on the first day of data collection. The physiotherapists provided individual verbal incentive 1) elbow flexion – "bend your elbow, bend, bend" and 2) knee flexion – "bend your knee, bend, bend". Participants were also instructed to keep their wrist in a neutral position (for elbow flexion), or for knee flexion – keep the thigh resting on the plinth. If the participant were struggling to maintain any instructions, the test would be repeated.

Table 6.1 Testing positions for elbow and knee flexion

Activity	Position	ASSA
Elbow flexion	Sitting on a chair, the arm to be tested resting on a plinth tilted to allow the whole arm to rest at around 45° with the floor.	The tester directed ASSA's contact area to the forearm's anterior distal area using the mark done before as a reference for the application of the device. The tester was asked to maintain a perpendicular orientation of the device throughout the movement.
Knee Flexion	Prone on a plinth with both feet dangling. Both arms were lying on the plinth but not holding onto it. Participant allowed 5°-10° of knee flexion before resistance is applied.	The tester directed ASSA's contact area to the posterior lower third of the tibia just above the tibial malleoli while maintaining a perpendicular orientation with the lower leg segment, maintained throughout the movement. Therapist facing the subject's feet. Test: Patient flexes knee while maintaining upper thigh always in contact with plinth.

For elbow flexion, the testing position was the same as in Chapter 5, and for knee flexion, the testing position was the same as in Chapter 4.

6.3.6   Data sonification

Sonification must use a data source to create systematic, relevant, and perceivable sound to its user/s. Therefore, using the HHD's input provides this research with a viable and reliable input to create live audio-feedback – figure 6.1.

Simplicity was fundamental in the development of sonification to minimise cognitive effort. In initial trials, a combination of audio cues using simultaneous data (instead of only one data stream) from angle and force as well as velocity and force were also used. However, the output was confusing and not instantaneously perceived. For instance, when this was trialled in manual muscle tests, it provided the listener with an unclear piece of

information, mainly due to the short period needed for the task, plus the physical effort needed to resist the movement. This issue was considered to be particularly relevant in stronger joints such as the knee or hip. Therefore, this option was rejected and only one sound parameter (instead of multiple sound streams) was used for the MMT. More details are found below.

To create the sonification, the author considered prior work as identified in the scoping review (chapter 2) where several authors developed their own working hardware which was then adapted to provide sound-feedback, this is essential to provide a curated and specialized sound output for the task being tested. In-depth analysis of the papers selected in the scoping review shows authors commonly used model-based and mapping sonification even when testing sonification with non-experts, this was the case for physiotherapists, this choice was also important as similarly to most participants in the studies from the scoping review had never used sonification either. In this research, the physiotherapists had never used sonification for manual muscle tests – this meant creating a clear-cut audio-feedback that could be usable in a task with short duration. In general, several parameters could be modified using PyGame, namely: Pitch; Duration; Instrument type and Volume.

To use sonification, one has to select the variables "to be sonified", in this thesis, the main outcome to be analysed in this way is strength – therefore, force was selected as one of the variables to be sonified. Considering the HHD can gather information about time, force, angle and angular velocity, one had to be chosen. However, in the previous chapter, results from the angle of peak force were not highly reliable, which could put in doubt potential benefits from sonification, so this parameter was discarded for this work. On the other hand, angular velocity could provide information about the participants' movement that had never been reported before in HHDs. Therefore, angular velocity was chosen. Burns and Spanier (2005) showed that different joint speed changes the maximal force output in make break tests, considering that peak torque values also change with the velocity of testing on an IKD it was hypothesised that controlling the speed of MMT could affect reliability. With this in mind, the segment's speed and the force resisted by the tester were selected as two data streams for sonification testing.

The sonification model was developed in Python. Two mappings were created: Sound model 1 – Force sonification; Sound model 2 – Angular velocity sonification (due to the issues with the software and COVID-19 explained previously, this sound feedback was not tested).

Figure 6. 1 From manual muscle test to sonification
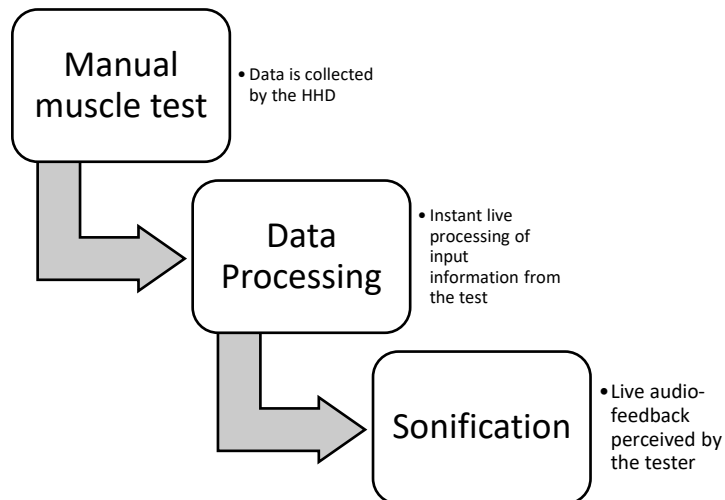
Designing the acoustic signal followed traditional and previously tested mappings in the literature, such as higher force or higher speed, would result in a higher tone in pitch (Dubus and Bresin, 2013). On the one hand, this would make it easier information for the user to decode the auditory information. On the other hand, when designing a sonification, it is critical to consider the listeners' expectation of what sound corresponds to the "target task" for a more congruent and effective task orientation (Ferguson and Brewster, 2018). In this case, this is particularly important, as higher pitch is commonly perceived and associated with higher energy (Dubus and Bresin, 2013). In theory, this should increase reliability by encouraging a maximisation of force production. It was also decided to use a discontinuous sonification formulation so that an increase in force production of 1kg could be undoubtedly perceived as such. If a continuous map had been chosen, the user might not be able to distinguish so clearly the correspondent increase in force.

The speed at which the segment moves seems to be one of the most important parameters to warrant a muscle test with high reliability and to maintain a constant lower speed seems

to correlate higher with the gold standard (see chapter 4). Therefore, the ability to control the speed at which the segment moves is essential to warrant reliability for peak force. Initially, it was hypothesised that the increase in pitch with the increase in speed would be a relevant model to test as this is in line with what previous authors found in sonification literature regarding kinetics and kinematics research (Dubus and Bresin, 2013). For that reason, it was brought forward as one of the final audio-feedback solutions.

Testing the best audio characteristics for the speed model-based sonification also involved the use of different durations of sound, from longer durations such as ½ of a second to shorter durations such as 1/30 of a second. The longer duration was initially considered, but it was difficult for a non-trained listener to discern pitch changes and act upon this change in actionable time, and therefore this was discarded.

Both models were developed to establish a clear and easy sonification that would allow immediate action, as it will be the case for both the experienced and inexperienced physiotherapists. With the above two main factors in mind – force and angular velocity - it was decided to use the MIDI notes from 21 to 108 which are the same notes one would listen from a Grand Piano ranging from A0 to C8 (27.50Hz to 4186Hz). This is a reasonably well-known sound and would not require learning from the participants when compared to more complex sounds.

6.3.6.1   *Sound model 1 – Force sonification*

As an initial development point, it was assumed that this model's auditory information needed to provide the physiotherapist with audio cues about the force applied by the subject. In theory, this should maximise manual resistance throughout the range of movement by incentivising the physiotherapist to apply more resistance. When testing this option, by trial and error, it became clear that either a higher note or a shorter period for each note would facilitate comprehension and enhance the results.

Firstly, to design this model, one had to consider the maximum force expected for the sample at hand. Previous data from chapters 4 and 5 indicated that a maximum force of

24kg was achieved for elbow flexion and 30 kg for knee flexion. Therefore, to encompass most individuals, the highest pitch was determined 5-10% above those levels to allow some latitude if a particularly strong individual was included in the sample. To establish the force model, force was linearly attributed to pitch which meant an increase in frequency corresponded to an increase in force applied.

For elbow flexion the force data (1 to 26kg) was directly attributed to a pitch note ranging from note 1 (27.50Hz) up to a maximum note of 108 (4186Hz), increasing by 5 MIDI notes for every increase of 1 kg, meaning a change in octave with every 3kg increase. Notes had the duration of $1/10^{th}$ of a second except for the tone played for forces less than 2kg. Loudness was also kept at the same level, regardless of force level.

Whereas in the knee flexion movement, the force data (1 to 32kg) was directly attributed to a pitch note ranging from note 1 (27.50Hz) up to a maximum note 108 (4186Hz), increasing by 4 MIDI notes for every increase of 1 kg, (average change in one octave with every 3kg increase). Notes had the duration of $1/10^{th}$ of a second except for the tone played for forces less than 2kg. Similarly, to the elbow flexion model, loudness was also kept at the same level regardless of force level.

In both cases described above, it was decided to use a regular increase in notes (every 4 or 5 MIDI notes), to correspond to a "step" in force, which was deliberately created to provide the tester with the auditory perception of a marked increase. Whereas if the model was progressing on a note by note basis, it would not give such sharp feedback to the user. This meant that a linear connection between strength and sound would be created. The larger the resistance provided, the higher pitch would be perceived by the user.

6.3.6.2   *Sound model 2 – Angular velocity model*

The angular velocity model used the same sound map for both joints, in contrast with sound model 1 which used different sound-parameters for each joint. When developing the sonification for the angular velocity model, the main aim was to achieve a specific speed while the test was being performed. To do this, a specific speed for the manual muscle tests was selected.

Due to the results from the previous testing, the speed that allowed a closer comparison with the IKD was 30º/sec, which was then selected as the "target speed". However, when using the target speed of 30º/sec, it was too difficult to perform the MMT due to the effort in maintaining the patient's joint moving at a fixed speed. Therefore, a "target range" (instead of a "target speed") for the task's execution was selected. A "target range" between 20º/sec to 40º/sec of segment displacement was then defined, allowing some margin of error while providing an indication about an acceptable segment displacement through the manual muscle test.

In practical terms, speeds lower than 20 º/sec were mapped to note 21 (27.50Hz), - duration of $1/10^{th}$ of a second; higher than 40 º/sec were mapped to note 108 (4186Hz) - duration of $1/30^{th}$ of a second. If the segment was moving between 20-30 º/sec, the note played was a 67 (391.99Hz). A grand piano was used as the instrument (so it was similar to the sonification force model), while loudness was kept the same for all attributes.

Although this sound map was created and was ready to be used, at the time when this was to be initiated, COVID-19 stopped the data collection, which meant only sound model 1 was compared to the no-sound group.

6.4   Data processing

A customised RStudio (Version 1.2.5033 © 2009-2019) program was written to process the data. A low-pass fourth order Butterworth filter was designed to filter the data from the device (ASSA – 10Hz). Data were tested for normality using Shapiro-Wilks ($p > 0.05$). Due to the nature of the analysis where comparisons between day or tester were performed, if a data point was missing for a particular participant, then the correspondent data point for the following day/tester was also removed. Therefore, different numbers of participants are expected. Muscle strength is reported in kilograms (Kg), joint angles in degrees (°) and impulse in Nm*sec. Mean and standard deviations (SDs) are reported.
Data analysis

Intra-tester reliability and responsiveness: The degree of correlation as defined by Schrout and Fleiss (1979) is done comparing between three repetitions for each tester is presented as – ICC (2,1). As reported previously, an ICC over 0.90 is considered

excellent, between 0.50-0.90 as fair/moderate to good and below 0.5 as poor. SEM and MDC are reported as in chapter 4 and 5.

Inter-tester reliability and responsiveness: The degree of correlation between testers as defined by Schrout and Fleiss (1979) is done using the mean (force, angle or angular impulse) of three repetitions of all the maximal strength tests using ASSA and was calculated using intraclass correlation coefficient (ICC(2,3)). SEM and MDC were also assessed in line with previous chapters (chapter 4 and 5).

Differences between the control group and sonification: A dependent t-test was performed to compare the mean force and angular impulse production for each joint. Effect sizes are reported with Cohen's D with small effect reported as r = 0.10, medium effect as r = 0.30 and large effects as r = 0.50. If data was not normal, the correspondent non-parametric Wilcoxon signed ranked test was performed. Bonferroni correction was used when performing analysis for multiple comparisons with a p-value of 0.025.

6.5 Results

A total of 24 healthy participants were recruited, but one did not attend the data collection stages, in total, 23 participants were involved in this research. From 23 participants, 19 were females 4 males, the mean (SD) age of 22.65 years (2.90 years), and the ages range from 20 to 32. Mean (SD) height was 165cm (8.2cm), range 153cm to 185cm while mean (SD) weight was 63.10kg (7.16kg), range from 45.6kg to 85.3kg. Further analysis is presented in the following sections.

6.5.1   Intrarater reliability

Elbow flexion peak force data for each of the testers show good to excellent reliability for all testers when assessing intraclass correlation coefficient – ICC (2,1) ≥ 0.86 (table 6.2). In terms of responsiveness, the SEM were between 6.46-12.46% and the MDC between (17.89-34.53%) for elbow flexion (p<0.001).

Results for ICC (2,1) angle of peak force (Table 6.2) for elbow flexion demonstrate moderate to good reliability for two testers and poor reliability < 0.5 ICC for two testers

(tester 3 and tester 4) (p<0.001). For the testers with ICC≥0.50 angle of peak force values for SEM were around 14% for both testers and MDC was 38.65% for tester 1 - 40.45% and tester 2 – 40.45%.

Angular impulse reliability for elbow flexion was excellent for three testers with only one tester displaying ICC lower than 0.75 (tester 3) (p<0.001). SEM and MDC values in percentage were between 7.75% to 13.70% and between 20.73% to 37.95% for MDC (table 6.2).

Table 6.2 Intra tester ICC(2,1) – Elbow flexion

PEAK FORCE	ICC	95% CI	P-VALUE	SEM (%)	MDC (%)	PARTICIPANTS
TESTER 1	0.93	0.87 - 0.97	<0.001	0.92(9.06)	2.56(25.09)	23
TESTER 2	0.87	0.74 - 0.94	<0.001	1.21(12.46)	3.36(34.53)	23
TESTER 3	0.86	0.74 - 0.93	<0.001	0.86(6.46)	2.37(17.89)	23
TESTER 4	0.95	0.89 - 0.98	<0.001	0.87(8.22)	2.41(22.76)	23
ANGLE PEAK FORCE	**ICC**	**95% CI**	**p-value**	**SEM (%)**	**MDC (%)**	**Participants**
TESTER 1	0.66	0.45 - 0.82	<0.001	13.35(13.95)	36.99(38.65)	23
TESTER 2	0.69	0.49 - 0.84	<0.001	11.60(14.60)	32.14(40.45)	23
TESTER 3	0.47	0.21 - 0.70	<0.001	NA	NA	23
TESTER 4	0.33	0.08 - 0.60	<0.001	NA	NA	23
ANGULAR IMPULSE	**ICC**	**95% CI**	**p-value**	**SEM (%)**	**MDC (%)**	**Participants**
TESTER 1	0.94	0.87-0.97	<0.001	4.44(7.48)	12.31(20.73)	23
TESTER 2	0.95	0.91-0.97	<0.001	4.43(7.75)	12.26(21.45)	23
TESTER 3	0.71	0.51-0.85	<0.001	4.75(13.70)	13.15(37.95)	23
TESTER 4	0.90	0.81-0.95	<0.001	4.73(11.93)	13.09(33.53)	23

Intra-tester ICC for knee flexion peak force results, for all testers, show good reliability (table 6.3) with all testers with values above 0.83 (p<0.001). For tester 4, one participant developed pain in the posterior thigh and was unable to complete the test. Therefore, the total number of participants for that tester is 22. For knee flexion, responsiveness was 6.46-7.47% (SEM) and 17.89-20.69% (MDC).

Data from the angle of peak force for knee flexion (table 6.3) shows moderate reliability for male testers (p<0.001). The two female testers showed poor reliability for detecting knee flexion angle of peak torque (ICC < 0.5). Whereas knee flexion SEM was 18.23% for tester 1 and 21.42% for tester 2, MDC was 50.49% for tester 1 and 59.33% for tester 2.

Lastly, the angular impulse for knee flexion (table 6.3) was excellent for tester 2 and 3 with tester 1 and 4 displaying moderate ICC, with values lower than 0.75 (p<0.001). For knee flexion, SEM ranged from 3.30% to 13.28% and in terms of responsiveness 9.15% to 36.77% for MDC.

Table 6.3 Intra tester ICC(2,1) – Knee flexion

PEAK FORCE	ICC	95% CI	P-VALUE	SEM (%)	MDC (%)	PARTICIPANTS
TESTER 1	0.84	0.72 – 0.92	<0.001	1.21(7.47)	3.36(20.69)	23
TESTER 2	0.87	0.76 – 0.94	<0.001	1.06(7.24)	2.93(20.05)	23
TESTER 3	0.86	0.74 – 0.93	<0.001	0.86(6.46)	2.37(17.89)	23
TESTER 4	0.86	0.75 – 0.94	<0.001	1.02(6.46)	2.82(17.90)	22
ANGLE PEAK FORCE	**ICC**	**95% CI**	**p-value**	**SEM (%)**	**MDC (%)**	**Participants**
TESTER 1	0.58	0.34 – 0.77	<0.001	10.45(18.23)	28.95(50.49)	23
TESTER 2	0.55	0.30 – 0.75	<0.001	10.47(21.42)	29(59.33)	23
TESTER 3	0.47	0.21 – 0.70	<0.001	NA	NA	23
TESTER 4	0.23	-0.02 – 0.52	0.04	NA	NA	22
ANGULAR IMPULSE	**ICC**	**95% CI**	**p-value**	**SEM (%)**	**MDC (%)**	**Participants**
TESTER 1	0.65	0.38-0.81	<0.001	7.84(10.26)	21.71(28.42)	23
TESTER 2	0.97	0.95-0.98	<0.001	4.89(6.47)	13.55(17.94)	23
TESTER 3	0.93	0.87-0.96	<0.001	1.63(3.30)	4.50(9.15)	23
TESTER 4	0.63	0.41-0.81	<0.001	7.04(13.28)	19.49(36.77)	22

6.5.2   Inter-rater reliability

The inter-rater reliability for elbow flexion peak force and knee flexion peak force was good (ICC > 0.75) for all testers (table 6.4) (p<0.001), SEM and MDC were found to be at 9.86% and 27.32% for the elbow movement, and 8.07% and 22.34% for knee flexion respectively. At the same time, the reliability for the angle of peak force for all testers (table 4) was poor for elbow flexion (0.34) and moderate for knee flexion (0.59), SEM was 16.63% and MDC was 46.05%. Data relating to angular impulse shows good elbow flexion reliability and moderate for knee flexion (0.50), where SEM was 12%, MDC 33.26% for elbow flexion and 17.96% and 49.76% for knee flexion.

Table 6.4 Inter-rater ICC(2,3)

4 TESTERS	ICC	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
PEAK FORCE – ELBOW F	0.89	0.64 – 0.96	<0.001	1.07(9.86)	2.97(27.32)	23
PEAK FORCE – KNEE F	0.79	0.52 – 0.91	<0.001	1.20(8.07)	3.33(22.34)	22
ANGLE PF – ELBOW F	0.34	-0.14 – 0.67	0.08	NA	NA	23
ANGLE PF – KNEE F	0.59	0.21 – 0.81	<0.001	8.67(16.63)	24.01(46.05)	22
A. IMPULSE – ELBOW F	0.86	0.73-0.93	<0.001	5.71(12)	15.81(33.26)	23
A. IMPULSE – KNEE F	0.50	0.05-0.77	<0.001	11.39(17.96)	31.56(49.76)	22

6.5.3 Sonification model 1 vs No sound

In this section, comparisons are presented for the participants (9) that were tested with both no sound condition and sonification model 1. Bonferroni corrections were applied.

Intra-tester reliability for peak force (table 6.5) data show that the reliability increased in the sound condition with all tester improving the stability of their results except for tester 1. Tester 1 still displayed excellent reliability (0.92) under the sonification intervention. On the other hand, intra-tester reliability for angle of peak force (table 6.6) data display decreased reliability in the sound condition with all testers worsening their performance. Results for elbow flexion impulse (table 6.7) show an increase in all testers' reliability under the sound condition except for tester 4 whose reliability did not change.

Table 6.5 Intra-tester reliability for the no sound and sound condition (ICC(2,1)) – Elbow flexion Peak force

TESTERS	ICC –NO SOUND	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.95	0.86-0.99	<0.001	0.77(7.93)	2.14(21.97)	9
T2	0.84	0.59-0.96	<0.001	1.47(15.21)	4.06(42.14)	9
T3	0.83	0.57-0.96	<0.001	0.91(7.07)	2.53(19.56)	9
T4	0.91	0.76 – 0.98	<0.001	1.16(11.97)	3.21(33.14)	9
TESTERS	ICC –sound	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.92	0.73 – 0.98	<0.001	1.06(11.19)	2.93(31.01)	9
T2	0.93	0.75 – 0.98	<0.001	1.15(10.69)	3.17(29.61)	9
T3	0.94	0.81 – 0.99	<0.001	0.99(9.27)	2.74(25.68)	9
T4	0.94	0.82 – 0.98	<0.001	1.14(9.40)	3.16(26.03)	9

Note: T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Table 6.6 Intra-tester reliability for the no sound and sound condition (IC(2,1)) – Elbow Angle of Peak force

TESTERS	ICC - NS	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.34	-0.04 – 0.75	0.04	NA	NA	9
T2	0.77	0.39 - 0.94	<0.001	8.77(13.44)	24.30(37.22)	9
T3	0.74	0.41 – 0.93	<0.001	11.91(20.22)	32.98(56.02)	9
T4	0.57	0.17 – 0.86	<0.01	11.90(13.36)	32.95(36.99)	9
TESTERS	ICC –S1	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.09	-0.26 – 0.60	0.31	NA	NA	9
T2	0.61	0.21 – 0.88	<0.01	10.05(12.07)	27.85(33.44)	9
T3	0.62	0.21 – 0.89	<0.01	15.80(15.52)	43.76(42.99)	9
T4	0.21	-0.07 – 0.64	0.09	NA	NA	9

Note: NS – No sound; Sound – S1; T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Table 6.7 Intra-tester reliability for the no sound and sound condition (ICC(2,1)) – Elbow flexion angular impulse

TESTERS	ICC – NS	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.80	0.41-0.94	<0.001	4.77(9.51)	13.20(26.35)	8
T2	0.92	0.80-0.98	<0.001	4.12(8.42)	11.42(23.32)	8
T3	0.85	0.60-0.95	<0.001	3.98(11.72)	11.02(32.48)	8
T4	0.92	0.77-0.97	<0.001	2.88(9.42)	7.97(26.11)	8
TESTERS	ICC – S1	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.92	0.80-0.92	<0.001	5.46(11.49)	15.13(31.82)	8
T2	0.96	0.90-0.99	<0.001	3.60(6.94)	9.98(19.21)	8
T3	0.90	0.69-0.97	<0.001	3.41(10.04)	9.44(27.82)	8
T4	0.92	0.45-0.94	<0.001	3.97(7.49)	11(20.74)	8

Note: NS – No sound; Sound – S1; T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Results for the comparison between sound vs no sound in knee flexion peak force, angle of peak force and angular impulse are presented in table 6.8, 6.9 and 6.10. Knee peak force reliability (table 6.8) was kept at similar levels for tester 1 and tester 2 under both sound and no sound conditions but varied dissimilarly for tester 3 – whose reliability fell from 0.83 to 0.68. Whereas tester 4 increased from 0.89 to 0.94.

The knee angle of peak force (table 6.9) demonstrates a decrease in reliability for both tester 1 and tester 2 when under the sound condition to poor reliability levels. Tester 3 and tester 4, however, display an increase in reliability to 0.95 for both testers.

The sonification showed changes in angular impulse for the knee flexion movement (table 6.10) in only one tester. It increased the reliability for tester 1 from 0.75 to 0.90 and remains almost unchanged for all the other testers - with ICC over 0.84 for the sound condition.

Table 6.8 Intra-tester reliability for the no sound and sound condition (ICC(2,1)) – Knee flexion peak force

TESTERS	ICC –NS	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.85	0.62-0.96	<0.001	1.09(6.33)	3.03(17.52)	9
T2	0.89	0.70-0.97	<0.001	1.04(7.84)	2.88(21.72)	9
T3	0.83	0.57-0.96	<0.001	0.91(7.05)	2.53(19.54)	9
T4	0.89	0.69 – 0.98	<0.001	0.97(6.47)	2.69(17.91)	8
TESTERS	ICC –S1	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.85	0.59 – 0.96	<0.001	1.45(9.06)	4.02(25.08)	8
T2	0.88	0.67 – 0.97	<0.001	1.43(9.46)	3.95(26.19)	8
T3	0.68	0.20 – 0.87	0.001	0.98(7.17)	2.71(19.85)	9
T4	0.94	0.83 – 0.98	<0.001	0.87(4.92)	2.40(13.64)	9

Note: NS – No sound; Sound – S1; T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Table 6.9 Intra-tester reliability for the no sound and sound condition (ICC(2,1) – Knee flexion angle peak force

TESTERS	ICC- NS	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.50	0.08 – 0.84	0.01	11.69(21.00)	32.38(58.18)	9
T2	0.54	0.13 – 0.84	<0.01	11.63(23.89)	32.20(66.17)	9
T3	0.74	0.41 – 0.93	<0.001	11.91(19.67)	32.98(54.48)	9
T4	-0.06	-0.34 – 0.50	0.6	NA	NA	8
TESTERS	ICC –S1	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.41	0.01 - 0.81	0.02	NA	NA	8
T2	0.33	-0.10 – 0.77	0.07	NA	NA	8
T3	0.95	0.87 – 0.99	<0.001	8.00(9.47)	22.15(26.24)	9
T4	0.95	0.86 – 0.99	<0.001	10.25(11.74)	28.40(32.52)	9

Note: NS – No sound; Sound – S1; T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Table 6.10 Intra-tester reliability for the no sound and sound condition (ICC(2,1)) – Knee flexion angular impulse

TESTERS	ICC –NO SOUND	95% CI	P-VALUE	SEM(%)	MDC(%)	PARTICIPANTS
T1	0.75	0.38-0.92	0.002	6.21(8.21)	17.19(22.74)	9
T2	0.97	0.93-0.99	<0.001	3.79(6.85)	10.50(18.98)	9
T3	0.93	0.82-0.98	<0.001	2.62(5.41)	7.25(14.97)	9
T4	0.82	0.56-0.94	0.001	3.75(7.98)	10.39(22.11)	9
TESTERS	ICC –sound	95% CI	p-value	SEM(%)	MDC(%)	Participants
T1	0.90	0.74-0.97	<0.001	4.82(8.22)	13.34(22.77)	8
T2	0.95	0.86-0.98	<0.001	4.37(8.22)	12.10(22.76)	8
T3	0.90	0.75-0.97	<0.001	3.57(8.97)	9.90(24.84)	9
T4	0.85	0.59-0.95	0.001	3.40(5.97)	9.43(16.54)	9

Note: T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Inter tester reliability comparing elbow and knee flexion (table 6.11) from the mean of three repetitions for all testers reveals that ICC levels increase under the sound condition with p<0.01 for all comparisons. While in table 6.12 for the angle of peak force sound decreases reliability in elbow flexion and increases reliability in knee flexion. The inter tester reliability for angular impulse also shows different influence under the sound condition with an increase in ICC with sound 1 for elbow flexion but a decrease in ICC for knee flexion (table 6.13).

Table 6.11 Inter tester ICC(2,3) Peak force sound vs no sound elbow and knee – all testers

	ICC	95% CI	P-VALUE	SEM(%)	MDC(%)	PART.
ELBOW FLEX. NS	0.89	0.59-0.98	<0.001	1.07(10.21)	2.97(28.29)	9
ELBOW FLEX. S1	0.95	0.85-0.99	<0.001	0.94(8.67)	2.61(24.00)	9
KNEE FLEX. NS	0.75	0.22-0.94	<0.01	1.33(9.11)	3.69(25.22)	8
KNEE FLEX. S1	0.86	0.54-0.97	<0.001	1.22(7.83)	3.38(21.70)	8

Note: Flex – Flexion; NS – No sound; S1 – Sound 1

Table 6.12 Inter tester ICC(2,3) sound vs no sound - Elbow and knee Angle peak force all testers

	ICC	95% CI	P-VALUE	SEM(%)	MDC(%)	PART.
ELBOW FLEX. NS	0.64	0.09 – 0.90	0.01	10.66(14.23)	29.53(39.42)	9
ELBOW FLEX. S1	0.20	-0.41 – 0.73	0.24	NA	NA	9
KNEE FLEX. NS	0.12	-0.28 – 0.66	0.29	NA	NA	8
KNEE FLEX. S1	0.57	-0.2-0.90	0.05	19.64(25.71)	54.41(71.20)	8

Note: Flex – Flexion; NS – No sound; S1 – Sound 1; Part. - Participants

Table 6. 13 Inter tester ICC(2,3) sound vs no sound – Elbow and Knee Angular impulse all testers

	ICC	95% CI	P-VALUE	SEM(%)	MDC(%)	PART.
ELBOW NO SOUND	0.64	0.23 - 0.87	<0.001	16.51(17.01)	45.73(47.13)	9
ELBOW SOUND	0.75	0.43 - 0.91	<0.001	19.49(17.74)	53.99(49.14)	9
KNEE NO SOUND	0.42	-0.04 - 0.77	0.06	NA	NA	8
KNEE SOUND	0.28	-0.37 – 0.73	0.21	NA	NA	8

Paired t-test data were compiled below to compare the elbow flexion and knee flexion manual muscle tests with and without sound-feedback for both peak force and angular impulse, but not for the angle of peak force due to low ICC.

Paired t-tests to compare the elbow flexion with and without sound-feedback are presented in table 6.14. The comparison between interventions for tester 4 produced a statistically significant increase in force under the sound condition (sound 1), with a t = 3.11 (p = 0.01) and a large effect size 0.74. A boxplot for the mean force by tester according to the sound condition is presented in figure 6.2.

In Table 6.15 results for the comparisons regarding knee flexion for all testers are present. Data was also plotted as a boxplot as figure 6.3. Tester 4 produced more force under the sound condition (sound 1), with a statistically significant difference in the score for the sound group t = 2.56 (p = 0.01) with a large effect of 0.71. The Wilcoxon signed-rank test for tester 3, (differences are not normally distributed) indicated that the median for the sound condition were statistically significantly higher (13.98) than the median for the no-sound condition (12.83) ranks. The effect size is also large for this hypothesis at 0.77.

Table 6.14 T test comparison for elbow flexion peak force (Kg)

TESTERS	FORCE – NS	FORCE – S1	T (95% CI)	P-VALUE	MEAN OF DIFFERENCES	DF	EFFECT SIZE
T1	9.69 (3.39)	9.53 (3.67)	-0.30 (-1.34 - 1.03)	0.77	-0.15	8	NA
T2	9.66 (3.47)	10.39 (4.26)	1.38 (-0.50 – 1.97)	0.21	0.73	8	NA
T3	12.93 (2.07)	10.52 (3.99)	-2.48 (-4.65 - -0.17)	0.04	-2.41	8	NA
T4	9.7 (3.74)	12.01 (4.56)	3.11 (0.60 - 4.02)	0.01	2.31	8	0.74

Note: T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Figure 6. 2 Boxplot of mean force by tester comparing sound condition 1 with no sound – Elbow flexion
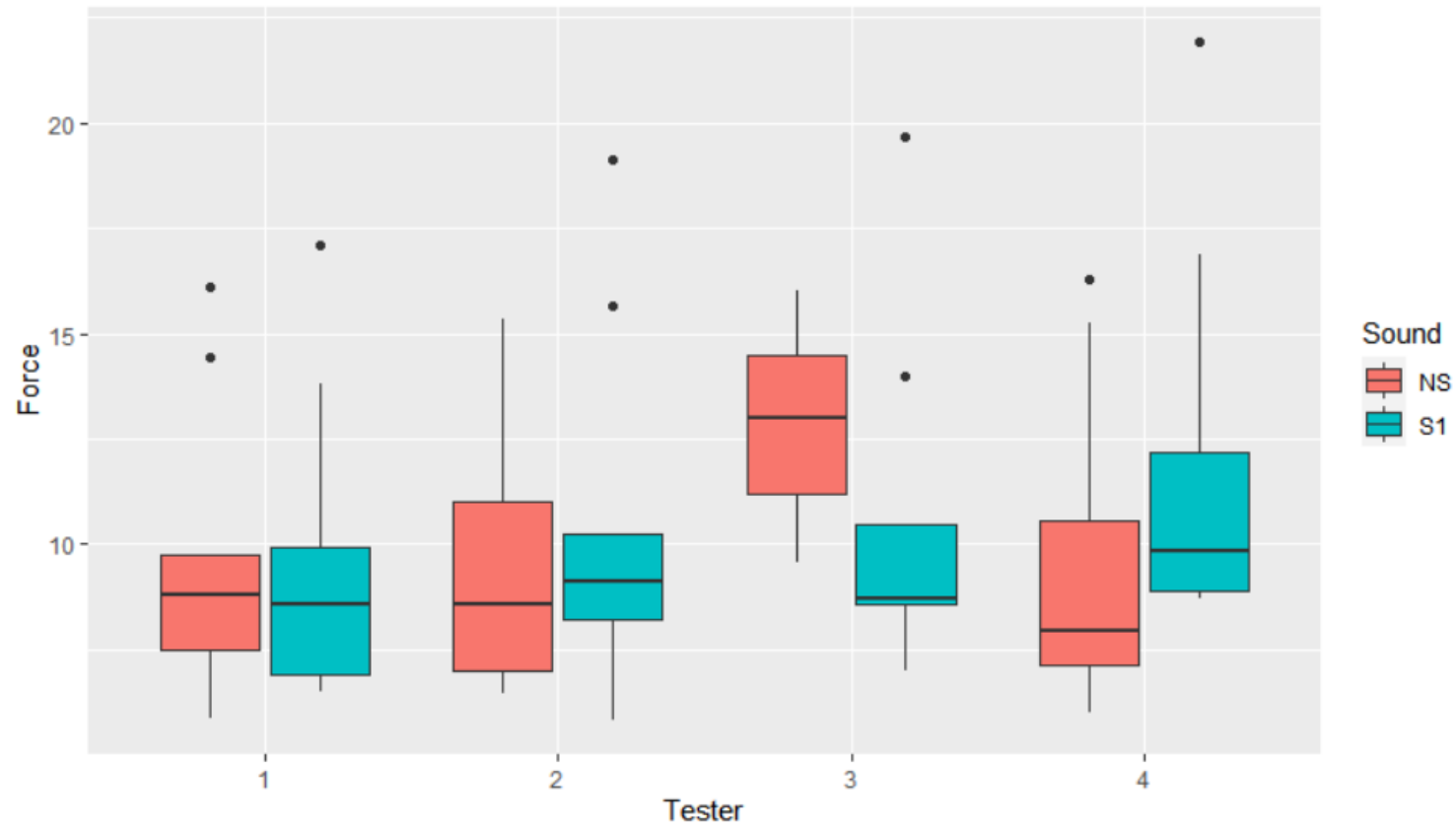
Note: NS – No sound, S1 – Sound 1

Table 6.15 T test comparison for knee flexion peak force (Kg)

TESTERS	FORCE – NS	FORCE – S1	T (95% CI)	P-VALUE	MEAN OF DIFFERENCES	DF	EFFECT SIZE
T1	17.78(2.82)	16.29(3.57)	-1.22 (-4.22-1.35)	0.26	-1.44	7	NA
T2	13.40(3.18)	14.99(3.94)	1.13 (-1.27-4.46)	0.23	1.59	7	NA
T3	12.83(2.13)	13.98(1.57)	Z= -2.17	0.02	1.15	7	0.77
T4	15.02(2.81)	17.59(3.70)	2.68 (0.30-4.83)	0.03	2.56	7	0.71

Note: T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4

Figure 6. 3 Boxplot of mean force by tester comparing sound condition 1 with no sound – Knee flexion
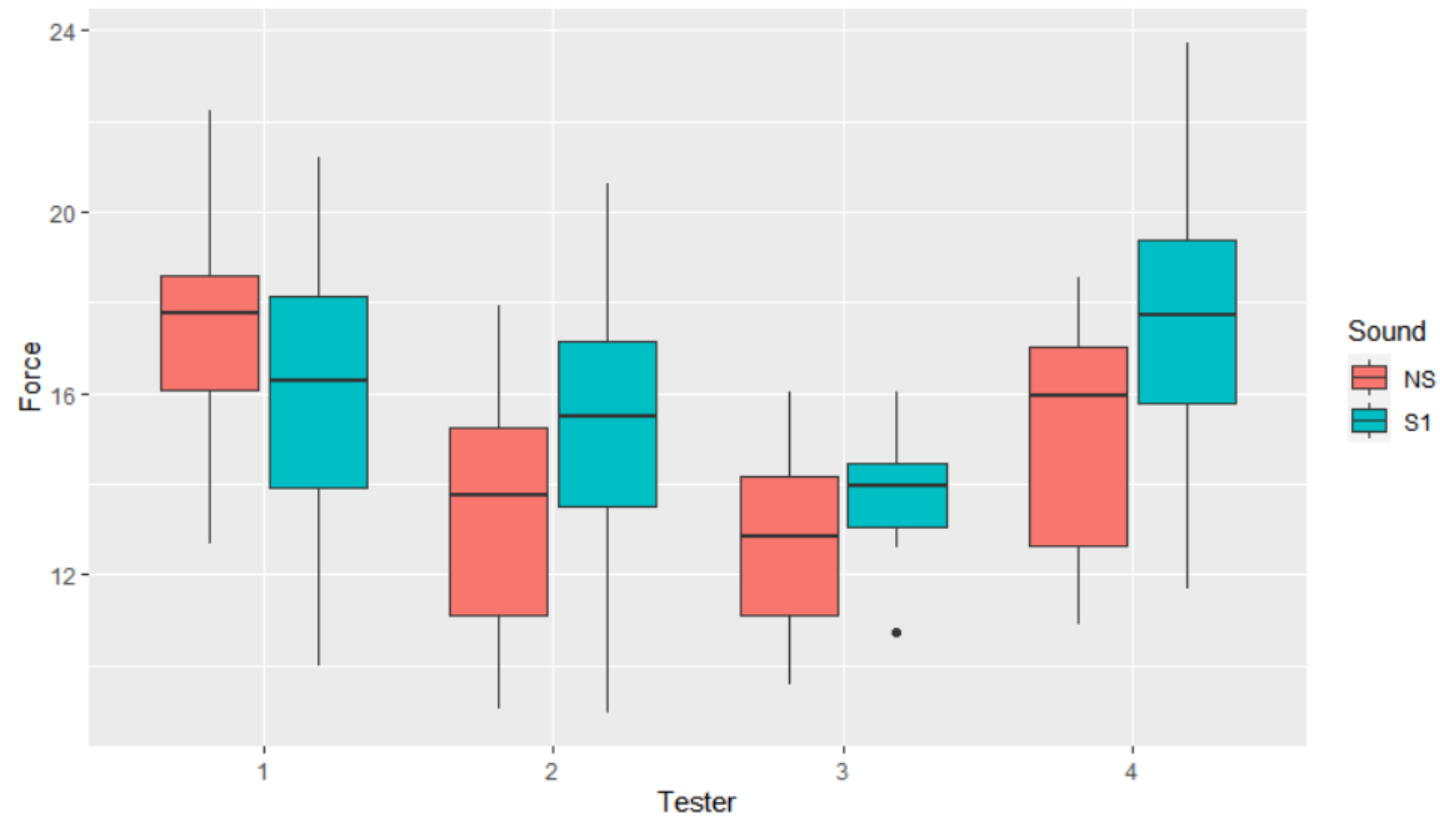
Note: NS – No sound, S1 – Sound 1

In table 6.16, there is a significant increase for angular impulse (p<0.001) in force (Mean = 52.90 Nm*sec) for tester 4 when using the sound-feedback for the elbow flexion movement, this corresponds to a large effect size (0.91). Elbow flexion angular impulse for the remaining testers was not statistically significant.

Table 6.16 T-test angular impulse mean elbow flexion (Nm*sec)

TESTERS	IMPULSE – NS	IMPULSE – S1	T (95% CI)	P-VALUE	MEAN OF DIFFERENCES	DF	EFFECT SIZE
T1	120.71(26.21)	112.89(48.19)	-0.58(-38-67 – 23.04)	0.58	-7.82	8	NA
T2	120.12(33.28)	121.31(44.69)	0.54(-24.10 – 38.95)	0.60	7.43	8	NA
T3	80.95(24.22)	71.42(22.91)	-0.10(-26.66 – 24.45)	0.92	-1.11	8	NA
T4	63.37(25.45)	117.50(35.13)	6.09(32.88 - 72.93)	<0.001	52.90	8	0.91

Note: T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4; DF – Degrees of freedom

Table 6.17 T-test angular impulse mean knee flexion (Nm*sec)

TESTERS	IMPULSE – NS	IMPULSE – S1	T (95% CI)	P-VALUE	MEAN OF DIFFERENCES	EFFECT SIZE
T1	232.99(40.82)	187.29(4.82)	-3.03(-81.37 - -10.02)	0.02	-45.69	0.75
T2	180.92(76.45)	172(66.46)	-0.31(-76.60 – 58.80)	0.76	-8.90	NA
T3	147.73(34.12)	128.45(33.23)	-1.15(-78.41 – 27.24)	0.29	-25.58	NA
T4	145.98(29.19)	178.27(26.38)	3.35(9.50-55.08)	0.01	32.29	0.78

Note: No Sound – NS; S1 - Sound 1; T1 – Tester 1; T2 – Tester 2; T3 – Tester 3; T4 – Tester 4;

The T-test for angular impulse, in table 6.17, shows there is a significant increase (p<0.025) in force (Mean = 32.29 Nm*sec) for tester 4 when using the sound-feedback for the knee flexion movement, this corresponds to a large effect size (0.78). However, tester 1 presents a change in the opposite direction with less angular impulse (-45.69Nm*sec) detected (p<0.05) and a large effect size 0.75. Knee flexion angular impulse for the remaining testers was not statistically significant.

6.6   Discussion

In this chapter, a group of physiotherapists and physiotherapy students with dissimilar experience in the use of hand-held dynamometers´ was assessed for reliability and potential effects of sonification on concentric manual muscle tests. Results are discussed below and compared with previous versions of the prototype.

6.6.1   Intra tester reliability and responsiveness in a group of experienced and inexperienced testers

Intra-tester reliability was only assessed when comparing three repetitions (ICC(2,1)) and not for the between-session reliability. For elbow flexion, intra-rater reliability for 23 participants shows that all testers were able to provide consistent values of maximal peak force good to excellent. This is similar to the high levels of reliability for elbow flexion found by Le-Ngoc and Janssen (2012) when testing elbow flexion movement using concentric movements. Knee flexion peak force data also exposed good reliability for all testers and is similar to results from chapter 4 for a previous version of the device in a similar testing position. In terms of absolute reliability and responsiveness, the SEM and the MDC for elbow flexion was 17-35% and knee flexion 17-21% respectively, these values are similar to the responsiveness obtained in the previous chapter, although they are expected to rise when performing testing on different days, as ICC results are usually worse when comparing data between days. More relevant data from inter-rater reliability is examined in section 6.5.2.

The angle of peak force reliability shows that male testers (one experienced and one novice) demonstrate higher reliability (moderate) compared with both female testers

(poor). These results for the female testers are similar to the data collected by Le-Ngoc and Janssen (2012), with only one physiotherapist performing the MMTs, which cast some doubt over the ability of these types of devices to be able to capture angle of peak force. The angle of peak force for knee flexion shows inferior results and increased variability across testers. While both male testers show moderate reliability, the female testers showed poor reliability. These results, accompanied by the high MDC values (above 38% for all movements tested), are similar to the previous chapter and do not support using the current dynamometer for the angle of peak force detection as they are set below the minimum ICC level of 0.75. In general, for the angle of peak force the male results were superior to the female testers; however, this might happen due to differences in individual strength which may play a role in keeping the device correctly orientated while performing the test, however, it was not assessed how individual strength differs between each tester. This should be investigated further, but considering the previous limitations in finding the angle of peak force accurately in chapter 5, the ability to detect it needs to be improved overall and neither biological sex nor experience seem to be the main limiting factor in achieving good results.

The angular impulse for elbow flexion results was moderate to good, with three testers able to consistently use the HHD with ICC ≥0.90 and only one displaying moderate reliability. The knee flexion movement results were excellent for two testers, but two other users displayed ICC values below 0.75 (one experienced and one inexperienced). Angular impulse has never been investigated before, and these results demonstrate they can be a reliable source when used in a concentric approach for some but not all testers. Tough the author hypothesizes that longer training sessions with the device where live result feedback is provided for angular impulse can improve the results, particularly considering that Le-Ngoc and Janssen (2012) were able to achieve acceptable results for total work in elbow flexion and knee extension and it was calculated in a similar manner to angular impulse.

SEM and MDC values in percentage are slightly lower for the angular impulse tests in the elbow and in knee flexion movements than on the previous chapter. This improvement is probably related to the type of ICC tested (between-session in chapter 5 and three repetition ICC in chapter 6). From these results, it appears that angular impulse data can

be reliably sourced with the concentric MMT using the newly developed HHD. Further emphasis needs to be given to adequate resistance given throughout the range of motion for this approach to work. This should be assessed in different joints to investigate responsiveness levels, which might further demonstrate the usefulness of angular impulse to measure strength.

6.6.2 Inter tester reliability and responsiveness in a group of experienced and inexperienced testers

Inter-tester reliability was one of the main outcomes of interest from this chapter; the testers' results seem to support the previous section's results with good interrater reliability. Good results were displayed in peak force and angular impulse data (except for knee flexion) but below standard, of ICC>0.75, for the angle of peak force in both joints. Inter-tester reliability for the knee flexion also increased as expected, compared to chapter 5, probably due to the increased stabilisation provided by the prone position, meaning this should be the position of preference for this movement, to the detriment of the sitting position.

Considering that in chapter 5, knee flexion and knee extension had similar inter-tester reliability, a similar effect - increase in reliability - would be expected if added stability was introduced to the participant's positions for the knee extension movement. This could be achieved in future studies by adding straps to the thigh or using a chair that allows for trunk stabilisation, albeit knee extension usually involves higher torque. Stabilisation has been discussed in previous research as an essential feature when using HHD´s which can decrease differences between measurements (Chamorro, et al., 2017). Nonetheless, this needs further confirmation regarding this device in future research.

Results for the four testers regarding the peak force interrater reliability are excellent for elbow and knee flexion, whereas the angle of peak force is poor for elbow flexion and moderate for knee flexion. This might be due to the device's test position and dimension, which might make it more challenging to maintain the device's alignment due to the participant's wrist closeness with the prototype (during testing, the body of the device is placed close to the participant's wrist) compared with the knee flexion movement where it is easier to avoid contact with the participant´s leg. These problems with excess

movement, which could impact results, were also reported by Le-Ngoc and Janssen (2012) but they did not assess this issue when comparing reliability for more than one tester.

SEM and MDC were in line with the previous chapter values, but higher for elbow flexion, which is likely to be due to a larger difference in absolute torque values produce due to the more heterogeneous sample of testers which might then increases standard deviation group values from which SEM is calculated, and the MDC derived. While SEM and MDC were lower for knee flexion movements, the author hypothesises this is due to the more stable prone position previously advocated in the last chapter. It is relevant to refer that there are no current guidelines for adequate responsiveness levels for commonly used dynamometers, but more importantly, they are unknown in MMT without HHD due to the ordinal characteristics of those measurements (Bohannon, 2019). According to the population and condition tested, responsiveness vary by joint and HHD and should be considered individually by researchers and clinicians.

Angular impulse responsiveness was similar to chapter 5 for elbow flexion, which again increases the interest in exploring this measure as a possible outcome of endurance/fatigue. However, it was worse in both reliability and responsiveness in chapter 6 for knee flexion, which means more emphasis should be given to training to improve this parameter in this type of user.

6.6.3 The effects of sonification in a group of experienced and inexperienced testers

6.6.3.1 Sonification effects in intra-tester reliability

The present study investigated the use of sonification in concentric manual muscle tests in a healthy group of participants. The results demonstrate that the tested sonification model (sound 1) increased ICC reliability for peak force in all testers but one (where the ICC was similar but slightly lower) when comparing intra-tester reliability for elbow flexion. Responsiveness, however, worsens for experienced testers when using sound 1 whereas it improves for inexperienced testers. This type of artefact when using sonification has been reported previously as it appears that some feedback might be detrimental or have no effect in experienced users due to their high levels of performance (Dubus and Bresin, 2015), however other authors point that differences between experts

and novices when using sonification are probably related to the type of task and sonification tested (O´Brien, et al., 2020). This also appears to be true in this research, with results differing when looking at different joints, which suggest that the tasks' specificity might influence the final results. For instance, when looking into the knee flexion results, the sonification appears to provide different results, with two testers maintaining similar reliability levels whereas one tester reliability improved (physiotherapy student) but another had worse results. Responsiveness for knee flexion only improves results for one tester (female and inexperienced) and worsens or does not affect the other three testers.

In sum, it appears that sound model 1 was consistent in improving responsiveness for one inexperienced tester in both joints, and it might therefore be useful as a learning tool for some individuals as reported before for sonification (Schaffert, et al., 2017; Maes, Lorenzoni and Six, 2019). This has not been reported before in HHD´s and is a new finding that implies that sonification can help some therapists obtain more accurate readings when using an HHD. It is also important to refer that this type of sonification was either indifferent or worsens their results for the remaining testers. As demonstrated previously in sonification, Dubus and Bresin (2015) also found different results across participants; further investigations are needed to understand which individuals benefit the most from it and how it can be optimized.

The angle of peak force in intra-rater reliability provides results depending on the joint analysed. Whereas sonification appears to worsen reliability in all testers in the elbow joint and also for two testers (male) in the knee joint, it improved the results from the two female testers to excellent reliability. This should be considered carefully as the sample consisted of only nine participants, considering that reliability is not consistently good across all testers and joints.

Angular impulse is not commonly reported in physiotherapy literature, but it should be considered in further research. The results from this section reveal that when using a concentric approach and an HHD with the ability to capture dynamic strength, it is possible to gather information about muscle performance through the range of movement, which should significantly impact functional rehabilitation. When using sonification model 1, results demonstrate good to excellent reliability for both joints tested, compared

to the no sound approach. In terms of responsiveness, there are also mixed results with the use of sonification, with improvements in SEM and MDC values for three testers in the sound condition for elbow flexion but with worse or similar results for knee flexion, again, and similarly to the peak force data, only the inexperienced and female testers improved angular impulse responsiveness in both joints when using the sound model 1, lifting the possibility of the ability of sound feedback to be used as a tool to enhance learning of MMT with the tested HHD. The use of sonification to support learning in complex tasks has received some support in the last few years and the results in this thesis appear to point in the same direction (Danna, et al., 2015; Jakus, et al., 2017; Schaffert, et al., 2017).

Lastly, it is important to refer that a high ICC(2,1) for 3 repetitions is less demanding to obtain than between-day reliability, which is a more insightful test for reliability and which this work has not investigated. Simultaneously, the sample size was small due to the technical issues described before and will need to be investigated further despite appealing initial results.

6.6.3.2   *Sonification effects in inter-tester reliability*

One of the major findings from the sonification testing is that peak force reliability is improved when using the sound model 1 in both movements tested, with changes from good to excellent in elbow flexion and moderate to good in knee flexion. Responsiveness is also improved in both joints by around 4% when using sonification. This is a novel finding in the area, which might indicate that the use of sonification can provide a decrease in testing errors if a large number of testers are needed (as it might be the case for research or if more than one physiotherapist). There is still a need to assess this finding in different joints to understand its implications for practice fully. This finding seems to be in line with previous research on the benefits of sonification from O´Brien, et al. (2020), although they suggest that providing information about task-related error might be more relevant for performance than sonification guidance which was tested in this chapter. Therefore, real-time error sonification should be tested in the future by assessing when comparing the reliability of MMT between testers of a different experience.

Data for inter-tester reliability was below 0.75 for the angle of peak force with and without sonification for elbow and knee flexion. The use of sonification decreased reliability in elbow flexion but increased it in knee flexion. The angular impulse results also show that sonification affects reliability differently according to which joint is tested. In elbow flexion, reliability was improved with sonification (to 0.75), but knee flexion reliability worsened with sonification (from 0.42 to 0.28).

In sum, sonification appears to improve inter-tester reliability and responsiveness in peak force but not for angle of peak force and angular impulse. Le-Ngoc and Janssen (2012) suggest that sound-feedback might be an interesting feature to develop in the future however, no future work was ever published to the author´s knowledge until now.

6.6.3.3 *Sonification effects in force production*

The t-test analysis shows a statistically significant increase (t=3.11 95%CI:0.60-4.02, p<0.01) in elbow flexion peak force when using sonification 1 with a large effect for the sample tested in one of the testers. Data from knee flexion also shows an increase in force production for two testers (tester 3: Z=-2.17, p<0.025; and tester 4: t=2.68, 95%CI: 0.30-4.83, p>0.05), both corresponding to a large effect size (>0.70). Other comparisons were not significant, and it appears that not all testers respond in the same way to the sound chosen in this research. It appears that less experienced testers might use the sound in a more productive way which should help improve results and achieve less discrepancy in comparison with experienced testers by increasing absolute force resistance and consequently higher values of peak force.

The differences arising from the effects in upper limb vs lower limb might have several sources which this thesis cannot fully unfold. It is hypothesized that the fact that resistance applied to the knee is more difficult to perform which could justify that two and not only one tester had statistically significant results in terms of force production. It is currently unknown how these results would fare when considering other muscular groups and researchers should investigate this dichotomy in the future with particular emphasis on flexor/extensor and upper limb vs lower limb.

The fact that some testers did not benefit from sonification has been found before in research (Sigrist, et al., 2016). Those authors explored the effects of sonification in users

with different levels of expertise and reported limited advantages in movement performance when investigating the effects of sonification in cycling technique which they related to short training time, tasks specificity or unclear sonification due to technical characteristics. In this thesis, the same issues could also justify the results with one major difference, MMT involved the interaction of two individuals, which can increase the difficulty in the creation of a playful and interactive sonification due to the dependence of force generation by an external component (the participant) and other factors such as fatigue.

In general, however, the effects of sonification appear to be emerging in several domains with overall evidence of the positive influence, being reported in interventions with different populations such as Parkinson's disease or stroke and tasks in sports as different as golf, swimming or rowing (Schaffert, et al., 2019).

Angular impulse results for elbow flexion also indicated a large effect for tester 4 (0.91), with more strength being produced across the range of movement (t=6.09, 95% CI:32.88-72.93), no statistically significant results were found for the remaining testers. There were also, significant changes found in data from the angular impulse in the knee flexion movement. Results suggest that while tester 1 (p<0.01) produced less force throughout the range of movement (table 6.17) when listening to the sonification. Whereas the novice tester 4 was able to increase force application in knee flexion (table 6.17). It seems that, at least for two users, the sonification was either indifferent (tester 2 and 3) or provided considerable aid in terms of angular impulse production (tester 4), meaning it might be able to guide concentric manual tests training in specific conditions. For the experienced user (tester 1) the audio guidance affected him differently with less force produced for one of the tasks assessed (knee flexion); it has been shown that experts and novices react differently to auditory information (Dyer, 2017a). This might have led to a decrease in force production throughout the range (angular impulse), meaning the sonification tested was not helpful in increasing angular impulse.

Lastly, although effect sizes were large, the sample for the sonification comparison and reliability test was small (9 participants), which means the advantages of using sonification should be further investigated as its benefits for clinical applications still need

to be made clearer. Further research is needed in specific populations and with different sonification mappings to explore uncovered relationships.

Different sonification approaches, namely, changing other variables such as loudness or instruments, were attempted in pre-trial in terms of the sonification model development. However, when attempting to use these during manual muscle tests, an increased reaction-time was noticed, probably due to sensory overload, rendering the sonification useless. The author does not argue that other approaches should not be trialled in future research, but considering the study design and aims, this was considered appropriate, and further expansion on the issue is welcome.

6.6.4   Strengths and limitations

The current chapter investigated for the first time how the use of sonification can support concentric MMT in both experienced and inexperienced physiotherapists. The research has also explored how testers with different experience and strength can display good intra and interrater reliability for both peak force and angular impulse. A concentric approach to MMT appears to be useful for both joints tested even if testers have less experience, highlighting the possible benefits of its use in clinical practice.

The main limitation of this chapter was that sample size was reduced due to COVID-19, this has an impact on the strength of the finding regarding the effects of sonification (although effects were large and statistically significant). Indeed, smaller samples will tend to find larger effects which can then affect the internal and external validity of this section (Button, et al., 2013). Researchers should bear in mind that the true effect is likely to be smaller than the reported here. In this text however, we did not opt to perform a post-hoc power analysis as it is unlikely it will produce relevant information besides what is already known – further participants were needed for adequate power (Althouse, 2021). Future research in the area can consider the effects obtained here to provide a rough estimate of effect which can then be used for future *a priori* power analysis.

Another limitation of the current work is that testers were not assessed for strength. This would have allowed comparisons between force resisted by strength level and contributed to provide a baseline reference for this type of HHD which could be helpful for future research.

Regarding missing data: Missing values on the last two data collections arose from pain occurring at the moment of testing which appear to indicate randomness and should not impact results. This was different from the pilot study´s missing data. In the pilot study data was missing as testers could not resist (inexperience/strength/not enough training) this was addressed and did not occur again in subsequent chapters. So as suggested by Kang (2013) a case wise deletion was the option, the sample size was also calculated with 20-30% more participants to account for this.

Lastly, in could be pointed that validity was not assessed when using sonification which can impact results. The use of sonification and reliability was done to assess if sonification changes that specific parameter. The same should be done for validity, but resources were not available for that at the time and a choice regarding which line of investigation to choose was made – it should certainly be considered in future research.

In hindsight, adjustments should be made to the study design in the future to prevent possible unwarranted effects on the no sound condition. For instance, the manual muscle tests should have taken place with the testers using earphones in both conditions (sound and no sound). This way, participants would not know if the tester was listening to the audio-feedback, or not, thus preventing them from potentially exerting more or less force.

The use of simple tones in the sonification model might be considered limited within the spectrum of augmented feedback for sonification. This can be improved in the next stages of research by using more instruments or other arrangements to enrich the sound experience and maximise sonification benefits. This point might be particularly important for a future speed sonification model as the planned range of sound-feedback might not be particularly useful as changes in speed of segment movement will impact peak torque production (see section 3.2). Authors researching the specificity of speed and its effects in MMT with the use of HHD should also consider the type at task being analysed as several previous authors have identified the stretch-shortening cycle as a major factor in force production which could impact results if an active stretch precedes testing as muscle elastic function influences force production (Miyaguchi and Demura, 2008; Turner and Jeffreys, 2010). More time for sonification development and participant training prior to the manual muscle test should also be taken into account to achieve this.

An important issue that could impact results whether in research or clinical practice is muscle temperature. The fact that temperature can affect muscle performance has been presented by researchers in the literature with both cold and heat being able to negatively impact muscle performance (de Ruiter, et al., 1999; Chaillou, et al., 2022), this can be of particular importance when assessing strength output using different warm-ups (Gogte, Srivastav and Miyaru, 2017); or in clinical practice when cold of heat interventions might be used before testing (Beaven, Kilduff and Cook, 2018); or if several trials increase muscle temperature above 39° (Brooks, et al., 1971). Future researchers should consider the implication of these issues in particular if several repetitions are demanded from participants and minimize differences between protocols while clinicians should be made aware of the differences that their interventions might have on the results of MMT.

6.7 Conclusion

This chapter shows that testers with different experience and gender are able to perform at acceptable and similar levels in terms of inter-tester peak force reliability and responsiveness when using concentric MMTs in elbow and knee flexion. More specifically, good reliability was also achieved for angular impulse in elbow flexion but not in knee flexion (inter-tester). Intra-tester reliability varies across testers but was generally moderate to excellent for peak force and angular impulse in both movements.

Sonification improved inter-tester reliability in both elbow and knee flexion for peak force. Intra-tester reliability was also improved in most tests for most physiotherapists, with one tester (inexperienced) showing increased force resisted in both movements when using the sound-feedback. These are novel discoveries that highlight the potential of sonification as a tool for concentric manual tests.

Manual muscle tests using a concentric approach are a reliable tool for a heterogeneous sample of physiotherapists with different degrees of experience for gathering peak force and angular impulse. Over the next chapter, a focus group to aid device development and deployment is reported.

7. ACCEPTABILITY AND IMPLEMENTATION – FOCUS GROUP WITH EXPERIENCED PHYSIOTHERAPISTS

7.1 Introduction

The main reason to develop a prototype is to create a new product that helps solve a specific problem. In any industry, but particularly in healthcare, prototype development should be accompanied by feedback from relevant stakeholders to optimise its function while minimizing design flaws. By assessing a device´s usability, one can gather information about how the device is perceived in its various facets, namely: ease of use, design issues, software problems, and other issues that might impact its acceptability by the end-user. Usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction, in a specific context of use" (International Organisation for Standardisation, 2018).

Schubert, Mühlstedt and Bullinger (2014) argue that device development should also be focused on how the device will be used in practical terms - in other words, the process is recommended to be human-centred. Hence, when developing new technology, it is of the utmost importance to prioritize user-focused approaches to avoid flawed interfaces. This will increase usability and prevent incidents that might provoke injuries to users (Imada, 1991; Garmer, Ylvén and Karlsson, 2004). Usability can be tested in several stages of product development through different approaches: expert-interviews; focus groups; questionnaires; user-interviews; and usability tests (Caplan, 1990; Nielsen, 1994; Schubert, Mühlstedt and Bullinger, 2014).

There is no broad tradition of performing focus groups in physiotherapy research, although they have been used before to support prototype development (Bruseberg and McDonagh-Philip, 2002; Zaina and Álvaro, 2015). Sim and Snell (1996) argue that focus groups have an advantage compared to other qualitative methods of data collection that can be important in specific scenarios, for example: for "respondents to have common experiences, interests and understanding", which is clearly the case for this group of experts who tested the device. This approach has also been used to gather data in the

development of a prototype to promote physical activity (Ganesan and Anthony, 2012) or to gather information about different interventions after lumbar discectomy (Rushton, et al., 2017).

During this research, the prototype (ASSA) idealised and created by this thesis author has only been tested in healthy participants and by a small group of physiotherapists, which means the sample size would be restricted when compared with similar research. For this reason, the author opted for a mini focus group approach as suggested by Kamberelis and Dimitriadis (2005), when only a small group of specialists is eligible for the focus group. The mini focus group aimed to allow the flow of the conversation and elicit new ideas and approaches to the assessment of muscle force and the development of the prototype; this was thought to be preferable to a one-on-one interview.

The main advantage of using a focus group when compared to other approaches is the interaction between group members and the interviewer that can facilitate the flowing of ideas and the production of data while gathering personal insight in a semi-structured and open environment (Breen, 2006). This might be particularly useful with less eloquent or shy participants; however, it is important to keep in mind that it can easily work in the opposite direction if the group has a very dominant participant or someone that is perceived as a hierarchic superior – as it was the case for the experienced testers and the physiotherapy students. This was why physiotherapy students who participated in chapter 6 as testers were not selected for the focus group.

One of the advantages of gathering information about different stakeholders at this moment in the project evolution is the fact that those inputs can be used to change the design and usability of the device considerably if needed, which would not be possible if the information was only gathered at a later stage in the project timeline.

7.2 Method

7.2.1 Design

The study is reported in line with the Consolidated Criteria for Reporting Qualitative Research (COREQ) (Tong, Sainsbury and Craig, 2007), further details are presented on appendix 10. One mini focus group was conducted in June 2020 was led by the primary researcher – JG - a male physiotherapist (Post-Graduate Researcher with previous

experience with qualitative research but not focus groups). The selection of subjects for the focus group was a convenience sample as all participants needed to have used the device previously. Physiotherapists that had used the device in chapter 5 and 6 were selected (a total of four). Participants were recruited by email and invited for an online meeting, all invited accepted to participate. Participants read a participant information sheet and consented to the focus group. The focus group was audio recorded using a dictaphone and then transcribed verbatim. No repeat interviews were performed.

The sample size consisted of four physiotherapists due to the reduced number of professionals who had trialled the device. This is in line with Kamberelis and Dimitriadis (2005) regarding mini focus groups, which are usually between two and five. Normally, a pilot study for the focus group would be done to gather feedback or change the planned questions, but due to the small sample, this was not performed.

All participants had time to express their views and add any relevant information to the notes throughout the process. The meeting was performed online through a video conference and only the researcher and participants were present. The audio was recorded, and the focus group lasted for around 1h30min. All questions were sent in advance to the participants to minimize this issue and can be found (the translated version to English) in Appendix 11 as suggested by Breen (2006). As recommended by Sim and Snell (1996), notes were taken during the conversation to gather information from non-verbal interaction - this was particularly important as no video was recorded due to privacy issues. Ethics approval was granted from the HSHPRC in Portugal (Appendix 12), from where the physiotherapists were recruited, and all participants provided written consent.

For this focus group, a thematic analysis following a phenomenological approach was used. Thematic analysis can be defined as "a method for identifying, analysing, and reporting patterns (themes) within data. It minimally organises and describes your data set in (rich) detail" (Braun and Clarke, 2006).

7.3 Data analysis

The audio recording was later translated to English by JG (fluent in both Portuguese and English) and transcribed verbatim; this was double-checked and sent to the participants (along with the audio recording) to guarantee translation integrity. Participants were assigned a two-letter name as a pseudonym. Thematic analysis using codes, which are assigned to phrases or groups of phrases, allowed the researcher to identify emergent themes (Breen, 2006).

The data analysis has one limitation, as it was performed by only one researcher, whereas it is recommended that at least two people perform the coding to minimize bias. To minimize this, codes and themes were sent to the participants afterwards in order to gather feedback regarding the themes approached. If the participants had new points to add, these would be included in the results – the participants suggested no changes.

7.4 Results

Participants consisted of 4 experienced physiotherapists (DC – male; MG – male; RM – male; SA - female) with 13 to 23 years of experience. Six themes arose from the analysis. The transcript and can be found in Appendix 11.

7.4.1   Theme 1 - Function

Participants indicated that they liked the prototype and that it served the purpose for what it was built for, while still being easy to use. The participants reported a positive overall view about the prototype: DC "In general I too also liked the prototype, even though it is a prototype and before the final version is ready it can be improved" similar input was provided by RM "I liked it, I liked it a lot. The concept of the prototype is an excellent idea and it can indeed bring added value if integrated in the physiotherapist's clinical practice. It's an interesting and useful concept", with RM adding that the prototype was easy to adapt to even for inexperienced users. MG and SA were also positive about the device, whit MG saying ", it looked to be very useful and it was easy to understand how to utilize it" and SA pointing "…I have a positive opinion of the prototype, even though

I have some questions that we will discuss next, but above all I have the impression that the prototype was well received by the participants…". DC added that its use was well accepted by the physiotherapy students whom were using the device –"(physiotherapy students) quickly adapted to what needed to be done and there were not any major hurdles - but also by the participants". Participants agreed that there were no extra constraints in the use of the device compared to a regular approach to manual muscle testing.

7.4.2   Theme 2 – Practical use of the prototype and sonification

RM and SA, both trialled the sonification mappings for force production. RM started by reporting that personally, he did not feel any effect but referred that "it could be beneficial for less experienced or recently graduated physiotherapists". While SA recalled that the physiotherapy student paired with her on the research reported positive feedback in terms of resistance application in the initial range of movement.

SA considered that the sound did, however, provide an "extra – like a positive input … like when you increase the music volume for a difficult task", which was something she related to her running. This is a compelling point for the use of sonification as it might help both therapist and patient keep a high interest in the movement even if fatigue or task difficulty increases (maybe during the last repetition of assessment or exercise).

DC explained that he did consider that sound is a positive feature as long as it can be turned on and off as it can serve different purposes, not only for assessment but also for live-feedback as part of the rehabilitation and exercise program: "I think the prototype can definitely make the strength evaluation more reliable, but I think the biggest advantage is when a physiotherapist wants to work with a patient within a certain, pre-defined, value of force in a specific range of motion...with this prototype I can use the sound frequency to say that I am working at an 80% 1RM and it is easier to make the patient achieve its goals than with a manual muscle test". DC also added "The prototype can do more than the traditional manual strength testing because it can do a dynamic evaluation – I am not constrained to one standard position… for instance it evaluates speed and this becomes a question not of where, but how."

RM also added input regarding the use of adequate load for strength training in physiotherapy which could benefit from the sonification – "…the sound feedback allows you to have a more rigorous control of the grade of strength applied to the manual test for each individual… you can now more easily adapt the strength training and take into account inter-subject differences". Later in the interview he also expanded this point further: "One of the issues we see sometimes is that people start the strength training by applying a certain load without really knowing why they are doing with that load. With a prototype that measures the force at an initial stage and that then allows you to apply that same strength along the range of motion, it is possible to make sure that during the movement we are always applying the ideal force for that muscle and specific angular position, and that is the value of the prototype. Not only assessment but also muscle training tool."

7.4.3   Theme 3 - Physiotherapy settings

The interviewed physiotherapists provided a myriad of settings where they believe the device would be useful and appear to agree that the device can be used transversely in rehabilitation. The participants started by drawing inferences from their personal experience:

RM - "I think it could be used in more acute situations… I am referring this with intensive care in mind, especially at the initial stages of strength recovery it can be particularly important. But I can also see advantages in any other phase within the hospital, where function gain is at its earliest stages … community based, when visiting a patient and where there are not a lot of available resources"

SA - "Naturally what comes to mind are the areas in which I worked ... I can visualize the use of prototype...in intensive care. But I can also see it being used in other areas, as it is quite versatile".

Another participant reported not only on a specific area of physiotherapy where the device could be used but also suggested that it could be used in a functional task:

DC - "It allows to evaluate specific situations, such as in the case of a stroke patient – how is the strength changing with the speed of execution – which gives me a lot of information on the quality of movements as simple as taking your hand to your mouth."

The same participant also referred that specific sports techniques can also be investigated: "If I consider an athlete with a key technical gesture, If I can break that gesture into more than one functional part, that can be evaluated with the prototype. Maybe a rugby tackle, the individual parts of the tackle…".

MG referred an important point with regards to his reservation on the use of these devices in an intensive care setting: "there can be limitations to the prototype there, since the patients can be intubated, with intravenous access, or there might be other electronic equipment that could interfere with this one…". The participant also added that the device could be used for hydrotherapy: "depending on how water-tight it is. I don't know anything like this that could be used in hydrotherapy" but also in sports "Anything that is related with research or high-performance, this prototype would be a perfect fit".

7.4.4 Theme 4 – Prototype concept and acceptability

Initially, MG voiced his reservations about the use of the prototype by some physiotherapists: "for the clinical practice of a physiotherapist that has to move from one place to another, I do not see it being used because even something as simple as clinical records is hard for some physiotherapists, so it would be the same with using this type of equipment". This led to a discussion with SA and DC whom both disagreed.

SA- "I believe the prototype will be easily accepted, it is simple, educational. Even with the participants, it was well-received straight away. There are some problems, as MG mentioned, like the wires and the weight. But that is more the characteristics, not the concept."

While DC pointed "In terms of concept, I think it is fantastic, and it will be very well received. Because if someone tries to sell me something that can give me specific data of what I am trying to evaluate and even other parameters that maybe I didn't even consider to evaluate in my clinical practice, also with an immediate registration of the data that can be accessed immediately and at any time, then that's something I'd want. And, more importantly, it does not change greatly what I already do in my clinical practice." He also

added: "… with this prototype, I simply have to use something between my hand and the patient, instead of using only my hand. So, with regards to the concept, the fact that I would only have to use something to hold between my hand and the patient and that can give me the right data means it will be accepted widely."

MG later clarified that he meant that people would not use the device if the device needed the laptop as it was tested in the data collection – "The concept I was describing included the computer, and what is needed now to use the prototype. But I see this as a great opportunity for many areas."

7.4.5   Theme 5 – Prototype implementation

The first barrier to the use of the device was its cost. DC – "There are certain constraints, like the price, and things that can be improved in the prototype - namely its functions and applicability." However, SA pointed that it is important to focus on the world market where focus is on the evidence-based intervention rather than focused on the short-term cost: "We need to contemplate other marketplaces, not just the Portuguese. If you consider the American market, if you need to present results of your evaluation, then people would quickly realise this can bring value. Portugal's reality is changing - when we say the patient will need 12 physiotherapy sessions, there's the question, why not 10 or 6, etc. there will be the need to show that what you are doing truly is working."

At this point MG pointed that the input from insurance companies can be a valuable force into introducing change for the use of the device but again, the price issue arose. RM – "The concept and the device are good. But no matter how useful it is, or how interesting it is and how important it is in terms of advancing scientific knowledge and clinical practice, at the end of the day what matters is the cost." SA felt this was reductionist and that pressures from regulators, peers and health care services also have an influence on how physiotherapists employ their money – "I think the issue is the price and the fact that you feel obliged to show results, these are connected. If there were more pressure to demonstrate results, people would consider the price differently ".

RM seemed to agree with this –" This will more easily get accepted into the market if it comes from the top, rather than from the bottom - if insurance companies start saying that this is the best practice, some clients will accept that and others will not". He also agreed

that clinicians need to clearly understand what the value is (monetary or otherwise) to their practice. "From the moment that people realise that a certain equipment costs x but will bring in the value of x plus something, that's fantastic, people will buy it. If people can't see this, they won't purchase the equipment, to put it simply. What DC said is important, if the equipment does not add any significant value to the clinical practice, people will not buy it." He also added an important point that is inherently connected to cost, the dichotomy between buying a product or a service and how it is maintained. "The other thing is maintenance. … So, it's different if you purchase a product that is yours and you own, or a product that you then have to pay an annual fee to maintain, at that point it becomes a service. I believe this prototype could be seen as a service".

The follow-up question asked the participants to reflect on the opinion already mention by RM, and provide information about what other problems could interfere with the implementation of the device. Participants seem to identify individual behaviour issues such as difficulty in adapting to changes and also the characteristics of the device.

SA – "…people will find it interesting, but after some time they will go back to what they do every day. They will keep doing what they did in the past, and that can be a big limitation in our profession…". One suggestion to facilitate the device´s acceptance would be to have experts with media exposure to use and talk about it. SA added: "Maybe marketing campaigns could help. Another thing is if we can convince some of the strategic key people that usually sell these ideas in physiotherapy to use it, that would be important."

Another individual aspect pointed out by the participants related to the difficulty in accepting change. As proposed by MG – "Another thing is the unknown factor, that can be relevant too, the willingness to try the prototype - some people will say they don't want to learn about the prototype because it will be seen as extra work when performing muscle testing, and they can still perform the test without it, grading it from 0 to 5 and doing it quite quickly. The time it takes to utilise it, can also be a factor, as with the prototype the physiotherapist would spend more time when compared to the traditional muscle testing. The device´s mobility can also be an issue, since it would mean I would have to carry one more thing."

Another participant brought up the importance of making sure the final product is a high-end, durable device: DC – "…make the prototype 100% functional, of extremely high quality, even if it makes it more expensive, it needs to be high quality … And also, if the equipment is really good, but is not very durable, that won't work either."

7.4.6   Theme 6 – Improvements

The interviewed physiotherapists shared a wide range of interesting ideas in terms of prototype development, emphasising hardware and some compelling points around software development.

7.4.6.1   *Hardware*

Ergonomics

All participants seem to agree that the size and shape of the device could be modified to facilitate handling.

SA –"…the way to hold the equipment, this needs to be improved…the prototype has to be easier to handle…".

DC – "When it comes to ergonomics, you need a handle that can be easily adapted to the physiotherapist's hand, like a ball handle or another type of handle."

RM – "… What I enjoyed the least was the ergonomics, the weight…".

All participants agreed that the device was heavy and that an improvement in that area would be important. DC – "…I would say the weight is still a problem as it is a heavy item."

Fall resistant

DC also pointed that these devices are moved quite often in clinical practice and might need to be carried by the user, thus increasing the risk of it being dropped and damaged. DC – "We are talking here about a piece of equipment that is portable and will be carried from one place to another, it will have to be quite resistant to falls as that might happen often…".

Contact cushion

Considerable importance was given to the prototype's contact patch with the patient and how that could increase adaptability to several segments, comfort, safety and hygiene.

RM – "The foam was perhaps too hard and did not bend or adapt to the patient's segment…maybe something like a gel type of material."

MG – "It could be something more bendable, that could adjust to the body part it was being used on. If, for instance, it was being used on the anterior lower part of the lower leg, some people will have a specific shape for which the prototype can become uncomfortable. "

SA – "We also need to consider that these days, whatever type of material has contact with the patient will need to be washable and easy to disinfect. "

DC – "For the part of the prototype that gets into contact with the patient, it should be something removable so it can be adapted to more than one area. You do not need a lot of different options, maybe one plain, one semi-round and one that would be rounder would be enough. If you have these adaptions, it is better in terms of safety".

Wireless

The device had a cable connected to a laptop for power and information exchange which the participants felt would hinder its practical use if not removed – all participants considered this an important point.

SA – "To me, the main constraint was the wire…Without the wire, the other concern would be the autonomy because if I need to use it but forgot to charge, then there needs to be a way to overcome that quickly; otherwise I might not feel inclined to use it again. "

DC – "Also, when it comes to the portability of the prototype, it should definitely not have the wires. … the prototype will need some kind of battery or perhaps even two, so you can use one while the other is charging because you do not want to make a patient wait for the equipment to be working. It is about trying to anticipate these kinds of problems."

An interesting point was also made about different areas of where the device might be used in the future and the fact that it could be developed to be waterproof. MG – "I will also add that the prototype should be watertight, maybe for its use in hydrotherapy as well. Not necessarily so that it can go underwater, but that can resist splashes, at the least. So, if there is an accident and someone spills something on it, it can at least not be damaged by that. "

7.4.6.2 Software

Software reliability

Occasionally during data collection, the testing would have to be stopped due to some technical issue that participants considered not to be acceptable for the final prototype.

DC – "What happened during our data collection was that sometimes we had to stop and restart it, and that cannot happen. RM was talking about the support and maintenance so people will not purchase something that will mean a change to their clinical practice, but that then has problems every week like a bug or something and needs to be repaired - that cannot happen."

Integration with other systems

Participants were also keen on the prototype's future version having the ability to connect or be able to send information to external systems, either for visualisation or analysis purposes.

RM – "It could be interesting to develop a wireless way to communicate not just with the laptop collecting the data but also other systems. Let us say I already have a data collection system in my clinic, I would like to be able to integrate this prototype… with software commonly used in physiotherapy, but also smartphones or smartwatches. "

MG – "One example is the kind of platforms that hospitals have, which means physiotherapists could register the data in those platforms and doctors would have access

to this data live when they had appointments with the patients. So, there would be a way of communicating between the doctors and the physiotherapists."

DC – "The software needs to give the possibility to export the data, not only to other types of software but also to clinical databases …It should be possible to send the graphs or data by email easily…ideally all done from an app.

The software only allowed the user on the laptop to visualize the data, and DC also thought this could be changed. DC – "…I believe it would be important to be able to have a way of displaying the information during the evaluation, perhaps with a very simple graph on a smartphone or even on the prototype itself"

SA – "As for the audio feedback, it would be important if it worked with any headset, not just some specific ones."

Easy to use interface

The usability factor was once again determined to be quite important for the software interface as physiotherapists have to be able to understand the device straight away.

MG – "Something valuable would be to develop software that it is easy to use and that anyone could easily pick up on. Even if you have no expertise with this technology. The prototype's portability and autonomy are something that needs to be improved, as that is very important these days.

7.5 Discussion

This is the first mini focus group to gather the perspectives of experienced physiotherapists on the newly developed prototype, which uses concentric movements for manual muscle tests. One of the advantages of the development of this prototype is that it was developed by an end-user (physiotherapist). Nonetheless, input from experienced professionals is part of the development of any new prototype, particularly in healthcare, and this focus group allowed the iterative process of device development to gather input from several individuals with extensive and varied clinical experience (Ferguson-Pell and Cardi, 1993; Tausch and Menold, 2016).

After performing the thematic analysis, six themes emerged regarding the prototype and its use: function; practical use of the prototype and sonification; physiotherapy settings; prototype concept and acceptability; prototype implementation; improvements (hardware and software).

Participants well received the prototype, and they felt that the concept was innovative from both the sonification and manual muscle testing perspective. They also reported that the physiotherapy students who trialled the device felt it was easy to use and could effortlessly adapt to the concentric test. RM – "To summarise, the concept is very good. One of the most positive factors is that it is quite intuitive to use, and that is important for clinicians' acceptance, as people do not have time to learn new things - the more intuitive it is, the better." Although the prototype is not the first of its kind, it is not common, and physiotherapists recognize its contribution to the improvement of the way manual muscle testing is currently performed.

The participants agreed that sonification can bring an important addition to manual muscle testing, particularly for inexperienced users, although those who tested it (SA and RM) reported not perceiving significant effects themselves, but that one of the inexperienced testers reported an increased perception in force resisted. It is important to refer that the participants were not aware of the data results from the sonification testing. Data from chapter 6 does show that sonification can support some testers in increasing reliability, particularly if they have less experience in both elbow flexion and knee flexion movements. Another important feature of sonification is the ability of the clinician to provide a quantitative force resistance to a training/rehabilitation program (Schaffert, et al., 2017). This means that the physiotherapist would be able to monitor the patient progress and adequately load the muscle with live-feedback data, which is not usually performed in strength training in rehabilitation. Sonification has, however, been used with promising effects in stroke patients with upper limb motor impairments, although only 15 patients were tested (Scholz, et al., 2016).

Physiotherapists were able to draw on their expertise and experience and provided relevant feedback on possible settings in which the device could be applied. In general, they conceived the device being used transversely in physiotherapy for strength testing and how it can help to improve current approaches. DC – "In terms of concept, I think it

is fantastic and it will be very well received… more importantly, it does not change greatly what I already do on my clinical practice…But with this prototype, I simply have to use something between my hand and the patient, instead of using only my hand". They hypothesized that the prototype could be useful in most settings, from community-based physiotherapy to sports and intensive care units, which is in line with the initial goal of this prototype development – to provide an easy to use, cheap and intuitive device for strength testing. Their opinion falls in line with previous research using isometric testing with hand-held dynamometers, which has been used in several fields: older adults (Arnold, et al., 2010); children with spina bifida (Mahony, et al., 2009); adults with interstitial lung disease (Dowman, et al., 2016); and intensive care units (Samosawala, Vaishali and Kalyana, 2016) among others.

Physiotherapists also pointed to the fact that the new type of assessment would allow an increment in current strength assessment due to its dynamic characteristics but also due to the ability to gather quantitative data from tested movements across the range of movement. MG pointed that "…where can an individual reach its maximum strength or if he should increase this strength at a different angular moment and that is a good example of where the prototype can be used...". While DC said – "…It also allows to dynamically evaluate and gather information about the *arc of movement* and not just one position. The prototype can do more than the traditional manual strength testing because it can do a dynamic evaluation - I am not constrained to one standard position, I can evaluate the strength applied in the execution of a certain movement, and not just the muscle strength in one task and that is a fantastic advantage."

DC and RM stated that they could recognize value in the use of sonification when using the device for training purposes (with a percentage of RM used as sound-feedback). DC pointed that: "More than being useful purely for the evaluation of strength, it's useful for the daily work with the patient. In my opinion this prototype can definitely make the strength evaluation more reliable, but the major advantage is for when a physiotherapist wants to work with a patient within a certain pre-defined value of force production within a specific range of motion…"; for RM – "…the sound feedback allows you to have a more rigorous control of the grade of strength applied to the manual resistance for each individual…a prototype that measures the force at an initial stage and that then allows you to apply that same strength along the range of motion, it is possible to make sure that

during the movement we are always applying the ideal force for that muscle and that specific angular position, and that is the value of the prototype. Not only assessment but also muscle training tool". Synchronisation between tester and patient was another possible clinical advantage of the use of ASSA with sonification as physiotherapists SA and RM agreeing but with DC pointing that sound could also have detrimental effects if it leads to less force being applied. SA also pointed that it could lead to increased motivation "..like when you increase a music volume for a difficult task, which I associated with running..".

After some discussion, it also became clear that they felt other physiotherapists would accept the prototype and its use of concentric manual muscle test because the concept is easy to grasp, relevant to current clinical issues and provides a novel answer to strength testing and monitoring. However, previous research has found that professionals can resist adopting new technological solutions (Rothgangel, et al., 2020). Despite this, the participants suggested several points that could facilitate the dissemination of the dynamometer. Similar issues/suggestion have been identified before - in research about technology acceptance - and are presented below paired with the suggestions from the physiotherapists, namely: price or other maintenance costs - which has been proposed as an issue for the implementation of these devices (Bohannon, 2019); involvement of insurance companies as a motivator for change – by providing cover in the use of eHealth (van der Meer, et al., 2020); peer pressure and policy by national health service – has been identified in the example of eHealth (Ross, et al., 2016); clear identifiable advantages for clinical practice and reliability of the device – these were similarly found in previous literature review about the implementation of electronic medical records (Boonstra and Broekhuis, 2010).

Lastly, physiotherapists were really committed to providing solutions to problems they faced when using the device with several suggestions on both hardware and software improvement. Regarding the hardware, participants suggested improvements on the device's shape, weight, and wire as the main issues. They suggested this needs to be changed for clinical use to match similar devices. While in terms of software development, the issues of data integration with other devices, intuitive interface and software reliability were pointed as their main concerns, most of these issues are in line

with previous research regarding eHealth technologies (Boonstra and Broekhuis, 2010; Rothgangel, et al., 2020).

7.5.1 Strengths and Limitations

This focus group has only assessed experienced physiotherapists, and therefore, they might have a limited vision on the potential issues of use and implementation of the device that could differ from novice physiotherapists. This will have to be investigated in future studies before full clinical deployment. At the same time, as this device was only tested in healthy participants, potential problems might arise with different populations, such as people with mental health issues or certain pain conditions or specific age groups that might struggle to adapt to new technology quickly. The future development of the device should also consider acquiring the CE Medical Devices label which could be needed for further testing in specific populations. This was never required by Anglia Ruskin University in any conversation with either the Ethics Committee or patent office, feedback obtained stated that it was not under the medical device label during the proof of concept phase while being used for research and not used in clinical populations. A future focus group with emphasis on the patient's perspective and advice would be beneficial.

The small sample size of this focus group makes it difficult to compare the findings with other research, particularly considering that prototype development is a particular niche within physiotherapy. There are, therefore, themes that might be unexplored, and this research probably did not achieve theoretical saturation as defined in previous literature (Breen, 2006). This might mean that the concerns presented by physiotherapists might not reflect all the issues of device usability.

Two physiotherapy students had previously used the prototype and could have been selected for the focus group, however previous research has identified that focus groups should consist of heterogenous groups of individuals and that individuals in different hierarchical positions (student/teacher as it was the case), should be carefully considered as it may hinder the contribution of those in lower hierarchical positions (Sim and Snell, 1996). It was also deemed, due to the experience of those testers, limited input could be given at this stage considering the goal of the focus group.

7.6 Conclusion

Strength assessment is considered central to the current paradigm of evidence-based practice due to the central role of exercise in rehabilitation. The ability to correctly identify, monitor and improve muscle deficits is one of the main issues in physiotherapy practice, and the device has the potential to address this, as was identified by this focus group participants. Findings from the focus group support the further development of the device by providing straightforward suggestions on which direction to take the device on and how to create a device that can be readily accepted in clinical practice. The prototype will be developed to integrate the recommendations from the clinicians. Nonetheless, to warrant further development, a medical engineer will have to join the project as expert input is necessary to adhere to all reglementary norms.

8. DISCUSSION

This thesis has described a new dynamometer's development and assessment to capture different muscle function variables in healthy participants while testing the ability of sonification to yield relevant results to clinical practice. The discussion will summarise key findings and original contribution to knowledge (in bold) while providing an overview of the developed work relative to current research.

8.1 Overview

The scoping review (chapter 2) explored how sonification has been developed with the potential to impact rehabilitation and motor control. The author identified critical areas of current research, the technology means used, and potential effects. Sonification has evolved rapidly in the last few years, but it is not generally used in physiotherapy. The scoping review demonstrated that most investigations in sonification used healthy participants. This is due to several reasons but mainly due to most research's exploratory nature in which investigators were testing new technological resources or approaches (Guerra, et al., 2020).

Sonification can only be helpful in physiotherapy if it can provide distinct gain to both therapist and patient. To explore sonification with this thesis meant approaching an issue in clinical practice that was both clear and wide-ranging. The use of MMT in strength assessment is one of those issues with relevance for physiotherapy practice (Bohannon, 2019). In theory, sonification could convey extra-feedback to improve MMT reliability, significantly impacting physiotherapists and patients. Strength assessment is widespread in practice, with previous research in Australian physiotherapists indicating it is used by over 70% of musculoskeletal practitioners (Abrams, et al., 2006), while using MMTs has been reported one of the most common interventions in physical medicine and rehabilitation for professionals working with amputees and neuromuscular conditions (Haigh, et al., 2001). Although no data exists in the author´s knowledge, HHDs are common but not as regularly used as they should be in physiotherapy, with previous authors suggesting it should be used more often due to its many advantages over manual muscle tests (Stark, et al., 2011; Mafi, et al., 2012).

Problems arise from the above issues that this thesis sought to improve: limited strength output from common HHD´s, the difficulty in performing reliable tests in stronger individuals due to the tester's strength limits, the restricted information from regular MMT due to its subjectiveness, the dynamometer's cost, and the absence of sonification integration in common HHDs. The author developed a new low-cost prototype and assessed concentric MMTs in terms of validity, reliability, and responsiveness to solve this issue. The need to develop a prototype arose from the technology identified from the literature review in which researchers used an interface for sonification; this meant being able to capture data streams at their origin to minimise lag. This case meant creating a novel technology that would provide the greatest freedom for sound software manipulation and sound design. However, before reaching that point, a new HHD had to be developed and tested.

The work presented in this thesis has its basis on the HHDs developed by Li, et al. (2006), Lintott, et al. (2007), Janssen and Le-Ngoc (2009) and Cadogan, et al. (2011). As presented in chapter 3 a new clinical application tool must be assessed for its reliability and validity. In summary, this thesis's results demonstrated that **the device is valid when compared to two different IKDs in both lower and upper limb movements**. The research also demonstrates that **the device can reliably gather peak force data in knee, elbow and shoulder movements when using both experienced and inexperienced testers.** On another indicator, **the device detected angular impulse reliably for some but not all testers in elbow flexion and knee flexion**. Lastly, the use of **sonification (tested in chapter 6) showed promising results in improving reliability and force production in MMTs with one inexperienced tester**.

8.2   Relative reliability

8.2.1   Intra-tester reliability

8.2.1.1   Peak force

In chapter 4, the results demonstrated that **the device -ASSA v2- was reliable for an experienced tester in between-session for peak force detection in knee flexion and extension**. Data for the **comparison between isometric and concentric tests also**

demonstrated that less force is produced at peak force with a concentric MMT and that reliability and responsiveness were also superior when using ASSA v2 with the concentric method. This is an important finding that means one can reliably use the concentric tests while reducing the problem of individual tester strength when testing joints that produce large torques and are more difficult to resist.

Following hardware and software improvements on the prototype, in chapter 5, the author investigated how the device would perform on upper and lower limb movements with a homogenous group of testers. Results revealed that **ASSA v3 was reliable when two experienced testers were using the device for knee extension and knee flexion** (see section 5.4.1.1). Whereas results from chapter 5 demonstrate that the device is still reliable, but relative values of ICC are lower than those obtained in chapter 4. The author hypothesised this was due to differences in stabilisation, as knee extension was tested in an IKD chair with the participant using straps (trunk and leg); whereas the knee flexion was tested in prone. The lack of stability of the segment tested has been pointed a limiting factor in obtaining reliable measures (Chamorro, et al., 2017). For future use, researchers and clinicians should use adequate stabilisation solutions when available to improve results.

ASSA v3 also demonstrated **good to excellent between-session intra-tester reliability in both shoulder abduction (both testers) and elbow flexion** (see section 5.5.1.1). Leggin, et al. (1996) also exhibited similar reliability when testing intra-rater reliability using a Nicholas MMT in shoulder abduction. Simultaneously, the inter-tester reliability was excellent in chapter 5 and good in Leggin, et al. (1996) for their two testers with the Nicholas MMT. It is relevant to notice the authors used, in both situations, ICC(3,1), which gives more optimistic results of reliability. More recently, Hirschmann, et al. (2010), investigated the use of a fixed HHD in several shoulder abduction positions for between-session reliability. Those authors' ICC levels were similar to ASSA v3 at 90° abduction but inferior in all other testing positions.

Interestingly their results were similar to ASSA when the trunk had no support, which is not usually recommended for HHD testing. Dollings, et al., (2012) also investigated shoulder strength reliability in standard clinical tests when using an HHD with ICC levels lower than ASSA. These are encouraging results **as the concentric approach for**

shoulder abduction is at least as reliable as commonly reported isometric approaches.

The interest in new low-cost dynamometers seems to be gaining traction. One of the latest approaches was the work from Romero-Franco, et al. (2019) using a low-cost dynamometer commonly found on online stores to test its reliability as a fixed dynamometer with excellent reliability in several isometric movements, including shoulder abduction and elbow flexion, however, they only tested 14 subjects. **ASSA obtained excellent intra-tester reliability (see chapter 5) with experienced physiotherapists and good to excellent intra-tester reliability with a mix of experienced and inexperienced testers (chapter 6)** while Romero-Franco, et al. (2019) do not state the experience of their testers. However, this is unlikely to be significant since they used a fixed dynamometer where tester's technique's importance should be minimal. The high level of reliability reported in this section is needed for appropriate clinical use and is similar to data from similar research with common dynamometers (Stark, et al., 2011; Schrama, et al. 2014).

8.2.1.2   *Angle of Peak force*

For ASSA v2 (chapter 4), the angle of peak torque results demonstrated **reliability lower than 0.75. This means its usefulness in clinical practice for knee extension and knee flexion movements might be limited**. Results were not consistently better in any of the movements tested in chapter 5, except for knee extension, where they were moderate for tester 2. The **poor to moderate reliability was maintained for all testers and both movements in chapter 6**.

The lower reliability results for the angle of peak force in intra-rater reliability suggest the concentric MMT might not be able to detect it, which might arise from several problems. Firstly, it might be inherently difficult to obtain this data due to the reliance on human resistance and ability. Previous research has demonstrated difficulty maintaining a horizontal orientation between the HHD and the tested segment, even in isometric testing (Ancillao, Rossi and Cappa, 2017). Secondly, individual participant characteristics also contribute to ICC values below 0.75, as even IKD protocols might struggle to have high levels of reliability for the angle of peak torque (Bernard, et al., 2012). From previous research, lower speeds of IKD testing appear to increase the

reliability for the angle of peak torque (Blacker, et al., 2010; Nugent, Snodgrass and Callister, 2015), but even though MMT was performed at a slow speed, results were still unsatisfactory. Although results are unsatisfactory, previous limitations from the use of the IKD have also been documented from previous authors (Wilhite, Cohen and Wilhite, 1992; Mayer, et al., 1994; Dauty and Rochcongar, 2001).

Lastly, physiological principles approached in section 3.2 might also contribute to this issue which are not related to the mechanical characteristics of the device but arise from changes in muscle output from length-tension and force-velocity relationship which could impact angle of peak torque. In particular tester-induced changes in terms of the resistance and speed of the movement tested which could have impacted torque production across the ROM and decrease reliability.

In future developments, it might be more interesting to use sensors that are external to the dynamometer itself, such as in the research by Li, et al., 2006 (but no data was reported for the reliability of angle of peak force) or by only using the inertial measurement unit sensor to detect ROM (Cadogan, et al., 2011). This might facilitate the obtention of both ROM and angle of peak torque information by minimising the inherent changes to the device's orientation in relation to the segment, which happened with the design presented in this work. The importance of using the angle of peak torque is still debated in the literature by some researchers, for instance, as a marker of a hamstring injury and re-injury (Timmins, et al., 2016) and should be carefully assessed if there is a real need for further development considering the resources, expertise needed and clinical relevance.

8.2.1.3 Angular impulse

This HHD's ability to capture angular impulse data is a novel contribution to knowledge that could bring interesting new insights into muscle function. The results from chapter 5 demonstrate that the device can gather angular impulse information. However, this is variable across joints. **Shoulder abduction and elbow flexion displayed a good level of reliability in both tester 1 and tester 2**, similarly to the results reported by Le-Ngoc and Janssen (2012) for elbow flexion tested in a similar position, although they reported total work. Whereas results were not as positive in knee extension as tester 1 was poor and good for tester 2, while the ICC for knee flexion was just below 0.75 for both testers. This is most likely related to the fact that there is not as much stabilisation in knee movements

compared with the shoulder and flexion activity, which should be taken into account in future research.

In chapter 6, as indicated previously, the author did not assess between-session reliability, but results from the **comparison of the three repetitions were excellent for three testers in the elbow flexion and excellent for two testers in the knee flexion movement**. The results from this thesis do not make it clear why there is a change in reliability values for different testers and different joints but considering that the new approach implies that testers need to apply maximal resistance to allow movement within the available ROM, it is likely that the technique used has a significant impact on the data from angular impulse. Total work, however, has been used and has demonstrated good results for Le-Ngoc and Janssen (2012), when investigating elbow flexion and knee extension and needs to be explored further to confirm these findings.

8.2.2   Inter-tester reliability

8.2.2.1   ***Peak force***

Results from Chapter 4 demonstrated poor reliability for knee flexion and knee extension. However, as explained in section 4.5.2.1 it is likely due to the tester's inexperience. This issue was addressed using testers with extensive MMT experience and a longer training process for testers to allow increased contact time with the device and the concentric MMT before the trial. This allowed for the improvement of inter-tester reliability in Chapter 5 with **ICC higher than 0.75 for shoulder abduction, elbow flexion and knee extension but not for knee flexion.** The lower value in ICC for the knee flexion is likely to have arisen from the testing position, which was chosen to mimic the IKD testing position and facilitate validation – and considering this result, it should not be recommended in future research using concentric muscle tests due to its increased stabilisation. However, this is divergent from isometric tests for knee flexors in the seminal work produced by Andrews, Thomas and Bohannon (1996). The change in testing position in Chapter 6, which increased leg stabilisation, led to improved ICC values in the knee flexion movement, a similar level of reliability was shown in chapter 4 when using a plinth for testing – if possible, the prone position should be used to increase reliability.

In conclusion, **inter-tester reliability is at an acceptable level for all the joints tested and can be used by several therapists, even with different levels of experience at the same time in healthy subjects** and is comparable to reliability obtained with other HHD´s (Dollings, et al., 2012; Romero-Franco, et al., 2019). The testers from chapter 6, although labelled inexperienced, had nonetheless more experience when compared to chapter 4 testers, as they were physiotherapy students who had undergone graduate training for both make and break tests, which should also justify the good reliability. This, however, means that it is unlikely that any person without previous strength training can provide adequate resistance when using the concentric method.

8.2.2.2 *Angle of peak force*

As expected, considering the reliability from the intra-tester reliability presented before, angle of peak torque results for inter-tester reliability from chapter 4 were also poor. Whereas Chapter 5 and Chapter 6, reliability was slightly better, but with results at lower values than 0.75. Nonetheless, future design changes might help transform this outcome into a reliable feature (see section 8.2.1.2).

8.2.2.3 *Angular impulse*

In Chapter 5, data **regarding angular impulse between all testers were moderate to excellent for shoulder abduction and elbow flexion** but below the acceptable level of 0.75 for knee extension and knee flexion. As suggested before (section 8.2.2.1), an increase in stabilisation for the knee extension and flexion could probably bring the angular impulse reliability above 0.75 for the experienced testers group. Whereas in **Chapter 6, results were good for elbow flexion and moderate (but below 0.75) for knee flexion**. The lower-than-expected result, despite stabilisation, for knee flexion might have happened from the difference in expertise which might suggest that although experience and strength do not appear to affect peak force (good reliability in chapter 6), they seem to affect angular impulse. The only similar research to this thesis, regarding angular impulse, was presented by Le-Ngoc and Janssen (2012) but only for intra-tester reliability. There is no similar research with published data regarding inter-tester reliability for the angular impulse, making comparisons impossible.

It is also essential to consider that testers might still be more concerned with finding a peak in strength, as it is their default technique, rather than the amount of strength produced across the range of motion. This should be considered in future training for testers if the angular impulse is to be used. Lastly, and if the angular impulse is to be used as a reliable tool, the range of movement and testing time should be standardised to improve testers' results. Sonification should be able to support these features due to its ability to convey several data streams simultaneously. One interesting feature that should be assessed in the future is power, however data available does not allow for current calculation of power, but data will be utilized from the findings of the present work to power further studies on this topic.

If ASSA demonstrates good reliability for angular impulse in different joints to a clinically acceptable level, its importance as an HHD could improve as no other HHD is able to do this. The ability to detect angular impulse could lead to its use in various fields where endurance tests are being investigated, which could be of relevance for physiotherapy. For instance, current research highlights the upcoming relevance of muscular endurance tests, which have been deemed reliable for shoulder abduction in twenty-eight healthy adults (Micheletti, et al., 2020); or in specific clinical situations, as demonstrated by O'Neill, Barry and Watson (2019) in their work which identified endurance deficits in patients with Achilles tendinopathy in a group of thirty-nine runners. Nonetheless, **its reliability for intra-tester and inter-tester reliability in shoulder abduction and elbow flexion is a new finding for new concentric MMTs**.

8.3 Absolute reliability and responsiveness

SEM and MDC values are closely related and derive from the ICC value for each joint. They allow clinicians to make an informed decision regarding the suitability of a specific measurement tool. When developing a new tool is also essential to compare with previous research to understand how it fares among its peers.

In chapter 4, **compared to an isometric test, the concentric movement produced smaller SEM and MDC, which is also a novel finding**. This means that when using this prototype for both tests (concentric and isometric), **the concentric approach is able to**

detect a smaller change in strength than comparable isometric tests in a healthy young population for knee extension and flexion. However, this was obtained with the prototype and not compared with other commercialised versions. Therefore it needs to be investigated further to confirm these findings.

Comparing with the different versions of the prototype for the same movements, **ASSA v2 (chapter 4) displayed a SEM=5.04% and MDC=14% for between-session reliability in knee extension and for knee flexion, SEM was 8.9% and MDC was 24.74%**. Whereas in the updated **version (v3), the knee extension SEM was 8.92% and MDC=24.72%**. This is in line with findings by Mentiplay, et al. (2015), who compared two HHDs in knee extension movements with a therapist who resisted HHD where SEM was around 9% and MDC close to 18%. However, this is not true for a fixed HHD, with Romero-Franco, et al. (2017) showing SEM of 10.6N while testing a new low-cost HHD, whereas ASSA's was 44.49N (4.59kg). It is worth noting that they only assessed 11 participants and did not explain how their SEM was calculated, but fixed dynamometry appears superior in joints that produce larger torques.

For **knee flexion, ICC was good (SEM=9.21%; MDC=25.50%)**, whereas Mentiplay, et al. (2015), found SEM, in a similar testing position, to be 7.40-12.53% and MDC 14.51-24.56% depending on if a Lafayette or Hoggan HHD was used. The worsening of results from chapter 5 compared to chapter 4 is likely due to the change in testing position for both knee extension and knee flexion. Whereas in the knee extension movement, the movement in the pilot test was performed in an IKD chair for increased stability, this was not available in chapter 5. In the knee flexion case, the movement was not performed in prone but in a sitting position to try and emulate the IKD position. However, both actions seem to decrease the reliability of the device, and the positions tested in chapter 4 should be assumed when possible. The return to the original testing position in chapter 6, shows increased performance for **inter-tester reliability where the prone position was used – the reliability increased to 0.79 and SEM and MDC also improved (reduced percentage) to 8.07% and 22.34%, respectively for ASSA v4**. Similar results were also reported when using a fixed dynamometer by two physiotherapy students with knee flexion SEM found to be 9% and MDC 24.8% (Thorborg, Bandholm, and Hölmich, 2013). Martin-San Augustin, et al. (2020) presented another example of research using a fixed HHD, where their SEM was below 3.3% for knee extension and 8.5% for knee

flexion; their values were lower than ASSA, probably due to the use of an IKD chair with straps for data collection. The fixed HHDs presented by Martin-San Augustin, et al. (2020) provided better responsiveness than ASSA; however, this prototype can also be used as a fixed dynamometer if and when practitioners deem it appropriate. Though, a fixed dynamometer is not able to gather information from a concentric muscle action.

Responsiveness (MDC) for shoulder and elbow movements were below 10%, below 25% for shoulder abduction, and 18% for elbow flexion in chapter 5, with similar results for inter-tester reliability in elbow flexion even when using inexperienced testers. Responsiveness for ASSA seems to be better than some research previously published, for instance, Dollings, et al. (2012), when testing their participants bilaterally using a Powertrack™ II Commander reported an SEM of 13.8-15.1N for shoulder abduction, wherein this work, SEM was 5.88N (0.60kg); and for elbow flexion, their SEM was 19.6-20.5N, and ASSA was able to retrieve an SEM of 7.06N (0.72kg). Nevertheless, they tested their participants on a treatment table with no back support, whereas the procedure for ASSA had back stabilisation for the shoulder test and arm stabilisation for the elbow flexion, which might justify the differences in favour of ASSA. The findings from Dollings, et al. (2012) are similar to research published by Romero-Franco, et al. (2019), who tested a fixed HHD and where shoulder abduction SEM was 14.2N and elbow flexion 12.9N, again **the results from ASSA are smaller for responsiveness which indicates it is able to find peak force data with a reduced acceptable error than isometric counterparts in shoulder abduction and elbow flexion.**

When considering the values obtained from the responsiveness tests, it is likely that in most settings the SEM and MDC obtained with ASSA are acceptable as it is for most other HHDs. When considering previous research about HHDs Kronborg, et al. (2017) investigated strength changes in post-op hip fracture patients and demonstrated an increase between 8-10% in only five sessions. Superior increases have been reported for stroke survivors (Hill, et al., 2012), where strength training has been shown to induce changes of over 75% in various muscle actions. Increases in strength in over 30% have also been documented in ACL rehabilitation programs. Hughes, et al. (2019) and Ishøi, et al. (2016) demonstrated in a randomised control trial that an increase in eccentric hip adduction could also lead to strength changes of over 35% over eight weeks. These well-documented muscle strength changes are an example of what a regular physiotherapy

rehabilitation programme can achieve (although not for every case) and how the concentric MMT performed with ASSA can be useful for clinicians, particularly its responsiveness levels.

Another important issue that closely relates to MDC is the Minimal Clinically Important Change Score (MCID), this has been described as *"the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"* by Jaeshke, Singer and Guyatt (1989). This concept has been further reinforced by a more recent paper by Cook (2008) where the author emphasizes that MCID encompasses changes that arise from clinical intervention and that are "meaningful for the patient". Several issues arise from the calculation of this concept, as it varies in regard to the outcome measure used, population, condition, among others.

The calculation can be derived from several methods, withs some authors even pointing SEM as an option (Wyrwich, et al., 1999) but it should be specific for the context where it is being investigated while using anchors such as functional outcome measures, with authors has recently as 2022 recommending authors to calculate their MCID for their research (Molino, et al., 2022). In sum, MCID should be established using several sources to create categories that are disease and population specific while addressing the patients´ perspective (Molino, et al., 2022; Mouelhi, et al., 2020). The author recommends that future HHD research is paired with patient reported outcome measures that could be used as anchors for MCID calculations as most therapists already use these in their clinical practice.

8.4 Validity

Results from the pilot test (chapter 4) demonstrated that the **prototype was valid when comparing to a gold standard in its ability to detect peak torque of concentric muscle action in knee flexion and knee extension**. However, the LOA for knee flexion and knee extension are too wide to allow one measure to be replaced by the other. Stark, et al. (2011) also shared this conclusion in a systematic review encompassing upper and lower limb research about dynamometry and concluded that the use of HHD is a valid measure

when compared to the IKD. Though they only provide results for ICC and correlation, further reviews should contemplate using BA analysis to clarify how different devices have a different bias in various joints. **In chapter 5, the results also demonstrate that the device is valid and can convey the same construct as the IKD,** but there is some bias (fixed and proportional – though this can be minimised in future data collections by using a calibration equation from the WLP regression) in every joint tested - that was not present with the ASSA v2. **The device also underestimates larger values of strength in knee movements.** On the other hand, **it overestimates shoulder abduction. The closest agreement between the IKD and ASSA was in elbow flexion, but similarly to the knee movements, ASSA tends to underestimate the "true" IKD value in stronger participants**, as it has been identified for other HHDs this issue is unlikely to be solved as even fixating HHD to compensate for the higher torque produced in larger joints seems to underestimate the output of an IKD (Martins, et al., 2017). This issue might arise for three different reasons. Firstly, the gravity effect calculated by the IKD might be different from the calculation performed to adjust for this issue (as presented before in chapter 4). Secondly, it might be owing to changes in torque related to moment arm. However, this was minimized by adopting similar testing positions. It is also important to note that the use of different attachments on the IKD which implies extra force to move the device could imply a different moment arm and consequently larger differences in the abduction movement. Thirdly, the force-velocity principle and its relationship with peak torque could provoke differences in the results that arise from the inability to control speed when using the HHD.

The validity results from the ASSA v4 appear to be slightly inferior to ASSA v3; several reasons are possible for this finding. Firstly, the participants were not strapped as in the pilot study (the pilot study used an isokinetic chair for knee extension concentric test and the prone position for knee flexion test). The chair and the prone position allowed for increasing segment stability which might justify these differences. Secondly, the worst results surged when comparing data from shoulder abduction, where participants reported increased difficulty in performing this movement in the IKD, with some reporting discomfort during testing, whereas when using ASSA, resistance was applied on the hand which might also have affected the results. Nonetheless, the approach followed in chapter 5 chapter is more in line with what most physiotherapists would have available in a

clinical setting (as most clinicians do not have an IKD chair available), which means comparisons between IKD and HHDs tend to differ more in terms of validity if segment stability is not warranted, as referenced before. The presence of adequate stability has been recommended previously as an important feature for the use of dynamometers (Chamorro, et al., 2017), and the test might need to be adapted in further research to allow for improved comparisons with the gold standard.

As stated before, comparisons with similar research in terms of LOA is difficult because of different methodological approaches or differences in reporting units (Martins, et al., 2017; Lesnak, et al., 2019). Authors in future research should aim to reduce these differences by using BA analysis for validity and report their findings for the LOA as a percentage of differences, as performed in this thesis, to prevent this issue.

One exception from validity analysis that allows for comparison with similar work is knee extension when compared with the IKD at 30º/sec (chapter 4 – section 4.4.3.1). Where ASSA underestimated IKD by -13.80Nm with a lower LOA of -82.19 and an upper LOA of 54.59. These are comparable with Hansen, et al. (2015) for a fixed HHD comparison with an IKD but their HHD overestimates the gold-standard by 52Nm and LOA ranges from -31 to 136Nm and to the work of Lesnak, et al. (2019) with their HHD also overestimating the IKD by 19.4Nm and LOA of $\pm$ 53.2Nm. In both their cases, HHDs, were deemed valid, but, as in ASSA's case, LOA were wide, making it challenging to ascertain what a determined result on the HHD equates to on an IKD, meaning one device cannot replace the other. Finally, as indicated before, other issues present at the time of testing might influence results such as the motivation of testers/participants and the verbal feedback provided (although this was standardized to minimize this) could influence validity (Schrama, et al., 2014), researchers should then consider that error will be compounded by the several variables at hand – dynamometer error, tester input, user participation among other already discussed.

Two features of ASSA have not been validated in this thesis, angle of peak force and angular impulse. Although both measures have been deemed reliable in certain circumstances or by previous authors, further work is needed to clarify their ability to integrate the clinical practice.

8.5 Sonification

For the first time when using HHDs, sonification was investigated. The basis of this work considers the effects postulated in previous research, which was reported in chapter 2, and expands on its use by exploring potential effects over a physiotherapist's evaluation process. The improvement in reliability and responsiveness was the main goal of the use of sonification for MMT. These effects were based, for instance, in the performance improvements reported by Effenberg (2016) using sonification as an adjunct to other types of feedback in rowing; by Schaffert, et al. (2017), who sonified force applied to the pedals of a Wattbike to aid movement training; or by Scholz, et al. (2015) which used sonification in stroke patients in upper movement tasks which lead to increase motor function.

This was supported by the findings using sonification in chapter 6, where external feedback appears to affect muscle test performance in different ways but improves performance in general. In intra-tester reliability for three repetitions ICC appears to increase the stability of the results with **three testers able to improve intra-rater reliability (ICC>0.90 for all testers) in both peak force and angular impulse in elbow flexion**. In knee flexion, the **sonification increased reliability, but in only one tester**, another lowered the ICC, and two remained stable in peak force. **Similar findings were reported in angular impulse with the two testers who displayed moderate reliability, improving when using sonification**.

Another **positive effect of sonification was the improvement in inter-tester reliability, which enhanced results for both joints tested**. This is another beneficial aspect of external feedback as sonification appears to be valuable in reducing differences between testers. This is relevant not only in research performed by different testers but also when different physiotherapists might assess the same patients in clinical settings.

Also, interestingly for clinical practice, peak force data for both elbow and flexion increased under the sound conditions. This was also the case for angular impulse in elbow flexion but not for knee flexion. In terms of mean force production, **sonification appears**

to contribute to an increase in peak force resisted in elbow flexion and knee flexion with a large and significant effect size in one inexperienced tester. **For angular impulse the sonification had the same effect for that tester, with large effect sizes in both joints, assisting in producing more force across the range of movement and at peak force levels**. Sonification appears to encourage force production in novice testers and minimize differences between testers. Thus, it might be useful as a tool for learning by providing external feedback, which has been shown to improve performance in novices (Wulf, 2013).

However, the sonification effect induced the male experienced tester to produce less resistance for knee flexion angular impulse. It is well known that experts and novices tend to perform differently when acting on the same task, with external feedback sometimes being detrimental in experts (Winkelman, Clark and Ryan, 2017; Couvillon and Fairbrother, 2018). This thesis's results also appear to support that hypothesis for at least one of the experienced testers, but more research is needed to show who benefits the most clearly.

As stated before, muscular endurance could be deemed a new outcome of HHDs. If this is to become a reality, then sonification can also be of use because an essential feature of endurance isokinetic testing is the sub-maximal effort required in this type of test. Sonification should be able to provide live-feedback about force exertion that can guide the task and improve performance as in similar research (Schaffert, et al., 2017).

Responsiveness levels were also improved when using sonification by around 4% in inter-tester reliability for peak force data, which suggest its ability to reduce measurement errors in clinical practice. However, no clear benefit could be found for the angle of peak force or angular impulse. There were also divergent results in terms of responsiveness from intra-tester reliability, with no clear trend.

In sum, ASSA appears to be superior to current MMT compared to common HHDs by increasing reliability and responsiveness. The advantages also appear to be magnified by the use of sonification, particularly in novice testers, while also decreasing measurement error in inter-tester reliability. Considering the criticisms that MMT has faced over the years, it is clear that physiotherapists need to change their manual test use. Research has shown they are limited in finding small to moderate changes in strength detectable by

236

HHDs (Hayes and Falconer, 1992; Peek, 2014; Nagatomi, et al., 2017; Pfister, et al., 2018). This thesis shows that a concentric MMT using the prototype developed can be relevant for physiotherapists in assessing strength with smaller SEM and MDC than standard HHDs.

8.6 Focus group

The focus group results were essential to obtain feedback to improve the current prototype from professionals with over ten years of experience. Their input contributed to providing indications about current prototype limitations and a clear path for the device's future development. Apart from the suggestions about hardware and software development, physiotherapists provided guidance about implementation strategies to facilitate the device´s dissemination which can be useful in future advances.

Physiotherapists were in general positive about the new prototype. They provided several recommendations to improve the current prototype, as mentioned in the focus group (chapter 7). Expert feedback is also an important step in device development before it is ready for commercialisation (Caplan, 1990). In general, physiotherapists had some reservations regarding the device's size and weight, its ability to withstand falls, the comfort of the patients, and wires' presence. These hardware changes are easily achievable and should be sought after in further development in order to compete with similar devices such as the MicroFET®2. Physiotherapists were also keen on providing further improvements on the software with the main recommendations provided: software reliability (which was an issue with some participants); the ease of integration with other commonly used systems (health records, mobile phones or others); and improved final user interface.

Feedback about sonification was positive, although the experts did not feel a clear benefit when using external feedback. Physiotherapists also felt that concentric MMT reproduces movements in a more functional manner which can also be seen as an advantage compared with current dynamometers. As there are some acute situations where a maximal isometric contraction is discouraged, concentric testing can provide a more functional overview of muscle performance. This also seems to be supported by

researchers who assessed a multi-joint dynamometer's reliability in older adults (Legg, et al., 2020).

The use of outcome measures by physiotherapists has been debated for many years, with several authors pointing to multi-level difficulties that impair practical implementation (Swinkels, et al., 2011). The group of physiotherapists also demonstrated ideas about issues that could impair the implementation of the device in clinical practice, such as cost, time consumption or individual resistance to change.

They also suggested facilitating clinical implementation, such as providing a high-quality device, a service that could accompany the device and provide further support to physiotherapists and the influence of health care services/insurance companies that can impact the introduction of new instruments like this as an outcome measure. Similar arguments were presented by several other authors regarding difficulties and suggestions to facilitate the implementation of new outcomes (Stevens and Beurskens, 2010; Swinkels, et al., 2011; McDonnell, et al., 2018). Physiotherapists also suggested that using physiotherapists with social media relevance and opinion-makers could increase the device´s visibility. This point has also been suggested in previous research to aid in implementing digital interventions (Ross, et al., 2018).

In sum, for the prototype to be accepted in clinical practice, its advantages must be apparent to both therapist and patient, with clear evidence available to compare it with other outcome measures. As Murray and Duncan (2012) suggested, in their systematic review about the use of outcome measures by allied health professionals, strategies to implement new outcome measures should be broad and impact different clinical practice levels such as "organisations, teams and individuals". The recommendations and reflections from physiotherapists shall be considered in further device development and when introducing the device to a broader audience.

8.7   Strengths, limitations and future research

This work's strengths include developing and testing a new prototype HHD for physiotherapy using appropriate reliability, validity, and responsiveness testing methods.

Another strong point was the use of different testers groups that attested its ability to be reliable regardless of experience and sex. For future sonification mappings in this field, it is also imperative to acknowledge that smaller segments and range of movement will also influence the type of sonification mapping and, therefore, might not be useful for every joint. Lastly, the use of a concentric method of MMT and the sonification is innovative and adds to the knowledge for strength assessment in physiotherapy by lifting the veil in terms of the potential use of concentric MMT and sonification assessment.

Methodologically, there are several strong points, some of them derived from the work of Scharma, et al. (2014). For instance, all testers were blinded to the results, and testing variance was also achieved by randomising the tester who assessed the participant first. The same authors also recommend that "authors investigating intra-examiner reliability of measurements in physical therapy practice should focus on reducing sources of measurement error to improve reliability", this is what has been done in this thesis by testing an alternative method to improve reliability and minimise measurement errors. To minimise bias, the author also blinded the participants to the sonification by having testers use earphones. However, this could be improved for future research by having testers use earphones regardless of sound presence.

Another strength of this research was that training for the use of HHD did not require individuals to obtain the same results of MMT in relation to another physiotherapist. Keep, et al., (2016) attempted to train testers so they could achieve a certain degree of accuracy for their measurements. The current protocol did not attempt to do that, as it would be improbable to occur in a clinical environment when testing several different patients. It can also be pointed out, that such results could contribute to a false sense of reliability, mainly as there is no literature documenting how long the ability to produce similar results subsists.

Limitations of this research arise from several points described in the following lines. One of the most important ones was the impact of COVID-19. Since the beginning of 2020, it has affected the whole world, and its impact was also felt in research. In this project, that meant the data collection for chapter 6 was halted due to an impending lockdown and no further progress was made on the effects of sonification. Although there was an initial thought that research could be resumed after the lockdown, this ended not being the case,

with several restrictions in place in Portugal, which rendered further data collection unviable. This has impacted the ability to draw more definite conclusions on the effect of sonification on reliability and responsiveness due to the reduced sample in chapter 6. The author recommends that more participants are used in future research to adequately power the sonification (see section 6.6.4 for further detail). Researchers should investigate the effect of sonification on the intra-tester reliability for elbow and knee flexion movements when comparing different sessions in terms of future direction. This type of reliability is more relevant clinically and more challenging to achieve than between repetition reliability. The use of sonification for HHDs should also explore how live sound-feedback can support patients in the assessment and as an exercise guide. Sonification might be able to incentivise patients and produce live feedback regarding the effort or guide exercise in terms of the range of motion, which can be useful to record work or angular impulse.

Due to the novelty of the device, there are only a few other papers that compare the use of concentric tests. Within those, sample size, testing procedures, and statistical approaches make it difficult to compare data and assess how the devices would fare in patients with disfunction or from different age groups. This leads to one of the caveats of this research: the results presented here were for a young and healthy population and are difficult to extrapolate to different clinical populations.

One specific point about the dynamic muscle testing that was not assessed and should be considered by future researchers is the speed at which the MMT is performed. This has been identified as a factor which could impact peak force and this was not controlled in the experimental chapters. This is closely related as to the reason why power was not able to be calculated, as this research did not standardize either the range of movement for each of the testing positions this would lead to asynchronous data where speed and range would interfere with the calculation for power between repetitions – this should be addressed in the future.

At the same time, ASSA does not fully solve the issue of HHDs where the tester struggles to resist higher torques; or is unable to maintain a perpendicular position of the device with the segment being tested. However, new fixed dynamometers have been utilized and have demonstrated high reliability and validity. Other factors will impact HHD results

such as patients/therapist motivation, fatigue, pain, HHD's/IKD position, ability to resist the movement and learning effects (Stark, et al., 2011; Kim, et al., 2014; Schrama, et al., 2014).

Future researchers should also investigate how future versions of ASSA fare compared to standard counterparts, as the only comparison from this thesis was the isometric test provided by the same device, which means findings for the isometric tests cannot be extrapolated to other devices. Validity should therefore be assessed by commonly available HHD´s.

In sum, fixed HHDs might be a good option for several settings but not all. It is also the case that for stronger joints such as the knee extensors, a fixed dynamometer is superior to ASSA. Still, not all settings/patients can be tested using a fixed dynamometer and therefore, an alternative such as the one presented here is still relevant as long as its limitations are known.

8.8   Implications to clinical practice

This thesis has found several implications to future clinical practice:

a) ASSA is able to reliably gather data when using a concentric MMT, with lower peak force values than the isometric test. This might be beneficial for weaker testers or when the maximal isometric voluntary contraction is not relevant or challenging to achieve.

b) ASSA using concentric MMT is reliable in recording angular impulse in shoulder abduction and elbow flexion, which could be developed into endurance tests and provide fatigue measures to clinicians. These have been used recently in research and are becoming essential parameters in rehabilitation that common dynamometers cannot measure.

c) The effects of sonification were statistically significant in one of the inexperienced testers, which showed an increase in peak force and angular impulse in both elbow flexion and knee flexion. It also seems to increase reliability between testers in both joints. The use of sound-feedback might support the technique by providing real time-feedback of force production, using auditory systems that can facilitate the perception of force production in terms of peak force and performance across the range of motion. This feature,

considering the large effect size in one inexperienced tester could be used in the future as a learning tool and could also help with the implementation of fatigue/endurance testing.

d) It is possible to create a reliable and valid low-cost HHD. The use of low-cost components means it can be built in most countries worldwide for less than £150, which is a considerable difference for currently available dynamometers. This issue is significant in countries outside western Europe and North America, where resources are scarce and the possibility of having a high-end dynamometer or an IKD is remote. The focus group participants also pointed out these issues (section 7.3.5) and should be considered when developing the technology to warrant low-cost production.

8.9 Conclusion

This thesis created a new tool from an early development stage to an innovative and low-cost solution. The use of the prototype HHD in clinical practice can improve manual muscle tests' intra-tester reliability, inter-tester reliability and responsiveness. These improvements were achieved using concentric MMTs with a low-cost HHD that can convey relevant kinematic information through sonification, specifically to inexperienced testers. At the same time, the reliability in capturing angular impulse opens the door for using this type of outcome in clinical practice. By approaching a common physiotherapy technique in a new perspective, this thesis provides a clear contribution to knowledge on both concentric manual muscle tests and sonification as an assessment adjunct in strength testing. Future research is vital to assess the benefits presented here compared to common dynamometers and specific clinical populations.

REFERENCES

Abrams, D., Davidson, M., Harrick, J., Harcourt, P., Zylinski, M. and Clancy, J., 2006. Monitoring the change: Current trends in outcome measure usage in physiotherapy. *Manual Therapy*, 11(1), pp.46–53.

Abu-Arafeh, A., Jordan, H. and Drummond, G., 2016. Reporting of method comparison studies: A review of advice, an assessment of current practice, and specific suggestions for future reports. *British Journal of Anaesthesia*, [online] 117(5), pp.569–575.

Aitkens, S., Lord, J., Bernauer, E., Fowler, W.M., Lieberman, J.S. and Berck, P., 1989. Relationship of manual muscle testing to objective strength measurements. *Muscle & Nerve*, 12(3), pp.173–177.

Alcazar, J., Csapo, R., Ara, I. and Alegre, L.M., 2019. On the shape of the force-velocity relationship in skeletal muscles: The linear, the hyperbolic, and the double-hyperbolic. *Frontiers in Physiology*, 10(JUN), pp.1–21.

Althouse, A.D., 2021. Post Hoc Power: Not Empowering, Just Misleading. *Journal of Surgical Research*, 259, pp.A3–A6.

Altman, D.G. and Bland, J.M., 1983. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician*, [online] 32(3), pp.307–317.

Amaral, G.M., Marinho, H.V.R., Ocarino, J.M., Silva, P.L.P., De Souza, T.R. and Fonseca, S.T., 2014. Muscular performance characterization in athletes: A new perspective on isokinetic variables. *Brazilian Journal of Physical Therapy*, 18(6), pp.521–529.

Ancillao, A., Rossi, S. and Cappa, P., 2017. Analysis of Knee Strength Measurements Performed by a Hand-Held Multicomponent Dynamometer and Optoelectronic System. *IEEE Transactions on Instrumentation and Measurement*, 66(1), pp.85–92.

Anderson, D.E., Madigan, M.L. and Nussbaum, M.A., 2007. Maximum voluntary joint torque as a function of joint angle and angular velocity: Model development and application to the lower limb. *Journal of Biomechanics*, 40(14), pp.3105–3113.

Andrews, A.W., Thomas, M.W. and Bohannon, R.W., 1996. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. *Physical Therapy*, 76(3), pp.248–259.

Anlauff, J., Cooperstock, J.R. and Fung, J., 2013. Augmented feedback for learning single-legged stance on a slackline. In: *2013 International Conference on Virtual Rehabilitation, ICVR 2013*. pp.162–163.

Arksey, H. and O'Malley, L., 2005. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, [online] 8(1), pp.19–32.

Arnold, C.M., Warkentin, K.D., Chilibeck, P.D. and Magnus, C.R.A., 2010. The Reliability and Validity of Handheld Dynamometry for the Measurement of Lower-Extremity Muscle Strength in Older Adults. *Journal of Strength and Conditioning Research*, [online] 24(3), pp.815–824.

Ayala, F., De Ste Croix, M., Sainz De Baranda, P. and Santonja, F., 2012. Absolute reliability of hamstring to quadriceps strength imbalance ratios calculated using peak torque, joint angle-specific torque and joint ROM-specific torque values. *International Journal of Sports Medicine*, 33(11), pp.909–916.

Baltzopoulos, B., King, M., Gleeson, P.N. and Croix, M.D.S., 2012. The BASES Expert Statement on Measurement of Muscle Strength with Isokinetic Dynamometry Produced on behalf of the British Association of Sport and Exercise Sciences by Prof Bill. *Sport Exercise Science*, (31), pp.12–13.

Barbalho, M., Gentil, P., Raiol, R., Del Vecchio, F., Ramirez-Campillo, R. and Coswig, V., 2018. High 1RM Tests Reproducibility and Validity are not Dependent on Training Experience, Muscle Group Tested or Strength Level in Older Women. *Sports*, 6(4), p.171.

Barbour, R.S., 2005. Focus Group Practice. *Sociological Research Online*, 10(1), pp.211–212.

Beasley, W.C., 1961. Quantitative muscle testing: principles and applications to research and clinical services. *Archives of physical medicine and rehabilitation*, [online] 42, pp.398–425.

Beattie, P., Isaacson, K., Riddle, D.L. and Rothstein, J.M., 1990. Validity of derived measurements of leg-length differences obtained by use of a tape measure. *Physical Therapy*, 70(3), pp.150–157.

Beaven, C.M., Kilduff, L.P. and Cook, C.J., 2018. Lower-limb passive heat maintenance combined with pre-cooling improves repeated sprint ability. *Frontiers in Physiology*, 9(AUG), pp.1–7.

Bernard, P.-L., Amato, M., Degache, F., Edouard, P., Ramdani, S., Blain, H., Calmels, P. and Codine, P., 2012. Reproducibility of the time to peak torque and the joint angle at peak torque on knee of young sportsmen on the isokinetic dynamometer. *Annals of Physical and Rehabilitation Medicine*, [online] 55(4), pp.241–251.

Bevilacqua, F., Segalen, M., Marchand-Pauvert, V., Peyre, I., Pradat-Diehl, P. and Roby-Brami, A., 2018. Exploring different movement sonification strategies for rehabilitation in clinical settings. In: *Proceedings of the 5th International Conference on Movement and Computing*. [online] New York, NY, USA: ACM.pp.1–6. Available at: <https://dl.acm.org/doi/10.1145/3212721.3212881>.

Biodex Medical System. System 3: structures, formulas and calculations. In: Biodex Medical System. Biodex System 3 advantage software: operations manual. New York: Biodex Medical System; 1991. pp. 1-99

Blacker, S.D., Fallowfield, J.L., Bilzon, J.L.J. and Willems, M.E.T., 2010. Within-day and between-days reproducibility of isokinetic parameters of knee, trunk and shoulder movements. *Isokinetics and Exercise Science*, 18(1), pp.45–55.

Bland, J. and Altman, D.G., 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, [online] 327(8476), pp.307–310.

Bland, J.M. and Altman, D.G., 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, [online] 8(2), pp.135–160.

Blazevich, A.J. and Babault, N., 2019. Post-activation Potentiation Versus Post-activation Performance Enhancement in Humans: Historical Perspective, Underlying Mechanisms, and Current Issues. *Frontiers in Physiology*, 10(November).

Bohannon, R.W., 1986. Test-Retest Reliability of Hand-Held Dynamometry During a Single Session of Strength Assessment. *Physical Therapy*, [online] 66(2), pp.206–209.

Bohannon, R.W., 1988. Make tests and break tests of elbow flexor muscle strength. *Physical Therapy*, 68(2), pp.193–194.

Bohannon, R.W., 1997. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. *Archives of Physical Medicine and Rehabilitation*, 78(1), pp.26–32.

Bohannon, R.W., 2001. Measuring Knee Extensor Muscle Strength. *American Journal of Physical Medicine & Rehabilitation*, [online] 80(1), pp.13–18.

Bohannon, R.W., 2005. Manual muscle testing: Does it meet the standards of an adequate screening test? *Clinical Rehabilitation*, 19(6), pp.662–667.

Bohannon, R.W., 2018. Reliability of manual muscle testing: A systematic review. *Isokinetics and Exercise Science*, 26(4), pp.245–252.

Bohannon, R.W., 2019. Considerations and Practical Options for Measuring Muscle Strength: A Narrative Review. *BioMed Research International*, [online] 2019, pp.1–10.

Bohannon, R.W. and Wikholm, J.B., 1992. Measurements of knee extension force obtained by two examiners of substantially different experience with a hand-held dynamometer. *Isokinetics and Exercise Science*, 2(1), pp.5–8.

Bohannon, R.W. and Corrigan, D., 2000. A broad range of forces is encompassed by the maximum manual muscle test grade of five. *Perceptual and Motor Skills*, 90(3), pp.747–750.

Bohn, D.A., 1988. Environmental effects on the speed of sound. *J. Audio Eng. Soc.*, 36(4), pp.223–231.

Boonstra, A. and Broekhuis, M., 2010. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Services Research*, [online] 10(1), p.231.

Bosquet, L., Gouadec, K., Berryman, N., Duclos, C., Gremeaux, V. and Croisier, J.L., 2015. Physiological Interpretation of the Slope during an Isokinetic Fatigue Test. *International Journal of Sports Medicine*, 36(8), pp.680–683.

Bosquet, L., Maquet, D., Forthomme, B., Nowak, N., Lehance, C. and Croisier, J.L., 2010. Effect of the lengthening of the protocol on the reliability of muscle fatigue indicators. *International Journal of Sports Medicine*, 31(2), pp.82–88.

Boyer, É.O., Bevilacqua, F., Susini, P. and Hanneton, S., 2017. Investigating three types of continuous auditory feedback in visuo-manual tracking. *Experimental Brain Research*, 235(3), pp.691–701.

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, [online] 3(2), pp.77–101.

Breen, R.L., 2006. A practical guide to focus-group research. *Journal of Geography in Higher Education*, 30(3), pp.463–475.

Brock, H., Schmitz, G., Baumann, J. and Effenberg, A.O., 2012. If motion sounds: Movement sonification based on inertial sensor data. *Procedia Engineering*, [online] 34, pp.556–561.

Brooks, G., Hittelman, K., Faulkner, J. and Beyer, R., 1971. Temperature, skeletal muscle mitochondrial functions, and oxygen debt. *American Journal of Physiology-Legacy Content*, [online] 220(4), pp.1053–1059. Available at: <https://www.physiology.org/doi/10.1152/ajplegacy.1971.220.4.1053>.

Brookshaw, M., Sexton, A. and McGibbon, C.A., 2020. Reliability and validity of a novel wearable device for measuring elbow strength. *Sensors (Switzerland)*, 20(12), pp.1–16.

Brown, L. ed., 2000. *Isokinetics in human performance*. Champaign, Ill: Human Kinetics.

Bruckner, H., Bartels, C. and Blume, H., 2011. PC-Based Real Time Sonification of Human Motion Captured by Inertial Sensors. In: *The 17th Annual Conference on Auditory Display (ICAD 2011)*. [online] Budapest, Hungary: International Community for Auditory Display.

Bruckner, H.-P., Theimer, W. and Blume, H., 2014. Real-time low latency movement sonification in stroke rehabilitation based on a mobile platform. In: *2014 IEEE International Conference on Consumer Electronics (ICCE)*. [online] IEEE.pp.264–265.

Bruseberg, A. and McDonagh-Philp, D., 2002. Focus groups to support the industrial/product designer: a review based on current literature and designers'

feedback. *Applied Ergonomics*, [online] 33(1), pp.27–38. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0003687001000539>.

Bui, K.-L., Mathur, S., Dechman, G., Maltais, F., Camp, P. and Saey, D., 2019. Fixed Handheld Dynamometry Provides Reliable and Valid Values for Quadriceps Isometric Strength in People With Chronic Obstructive Pulmonary Disease: A Multicenter Study. *Physical Therapy*, [online] 99(9), pp.1255–1267.

Bujang, M.A. and Baharum, N., 2017. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Archives of Orofacial Sciences*, 12(1), pp.1–11.

Burns, S.P. and Spanier, D.E., 2005. Break-technique handheld dynamometry: Relation between angular velocity and strength measurements. *Archives of Physical Medicine and Rehabilitation*, 86(7), pp.1420–1426.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J. and Munafò, M.R., 2013. Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), pp.365–376.

Cadogan, A., Laslett, M., Hing, W., McNair, P. and Williams, M., 2011. Reliability of a new hand-held dynamometer in measuring shoulder range of motion and strength. *Manual Therapy*, [online] 16(1), pp.97–101.

Caplan, S., 1990. Using focus group methodology for ergonomic design. *Ergonomics*, 33(5), pp.527–533.

Carter, R. E., Lubinsky, J., Domholdt, E., and Domholdt, E., 2011. *Rehabilitation research: Principles and applications*. St. Louis, Mo: Elsevier Saunders.

Chaillou, T., Treigyte, V., Mosely, S., Brazaitis, M., Venckunas, T. and Cheng, A.J., 2022. Functional Impact of Post-exercise Cooling and Heating on Recovery and Training Adaptations: Application to Resistance, Endurance, and Sprint Exercise. *Sports Medicine - Open*, [online] 8(1), p.37. Available at: <https://doi.org/10.1186/s40798-022-00428-9>.

Chamorro, C., Armijo-Olivo, S., De la Fuente, C., Fuentes, J. and Javier Chirosa, L., 2017. Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: systematic review and meta-analysis. *Open Medicine*, [online] 12(1), pp.359–375.

Chanpimol, S., Seamon, B., Hernandez, H., Harris-Love, M. and Blackman, M.R., 2017. Using Xbox kinect motion capture technology to improve clinical rehabilitation outcomes for balance and cardiovascular health in an individual with chronic TBI. *Archives of Physiotherapy*, [online] 7(1), p.6.

Chiari, L., Dozza, M., Cappello, A., Horak, F.B., Macellari, V. and Giansanti, D., 2005. Audio-biofeedback for balance improvement: An accelerometry-based system. *IEEE Transactions on Biomedical Engineering*, 52(12), pp.2108–2111.

Clark, R., Bryant, A., Culgan, J.P. and Hartley, B., 2005. The effects of eccentric hamstring strength training on dynamic jumping performance and isokinetic strength parameters: A pilot study on the implications for the prevention of hamstring injuries. *Physical Therapy in Sport*, 6(2), pp.67–73.

Clark, D., Dean, G., Bolton, S. and Beeson, B., 2020. Bench to bedside: The technology adoption pathway in healthcare. *Health and Technology*, [online] 10(2), pp.537–545.

Clarke, M.N., Ni Mhuircheartaigh, D.A., Walsh, G.M., Walsh, J.M. and Meldrum, D., 2011. Intra-tester and inter-tester reliability of the MicroFET 3 hand-held dynamometer. *Physiotherapy Practice and Research*, 32(1), pp.13–18.

Cohen, D.D., Zhao, B., Okwera, B., Matthews, M.J. and Delextrat, A., 2015. Angle-specific eccentric hamstring fatigue after simulated soccer. *International Journal of Sports Physiology and Performance*, 10(3), pp.325–331.

Cook, C.E., 2008. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *Journal of Manual & Manipulative Therapy*, 16(4), pp.82E-83E.

Coratella, G., Bellin, G., Beato, M. and Schena, F., 2015. Fatigue affects peak joint torque angle in hamstrings but not in quadriceps. *Journal of Sports Sciences*, 33(12), pp.1276–1282.

Couvillion, K.F. and Fairbrother, J.T., 2018. Expert and novice performers respond differently to attentional focus cues for speed jump roping. *Frontiers in Psychology*, 9(NOV), pp.1–9.

Cramer, J.T., Housh, T.J., Johnson, G.O., Weir, J.P., Beck, T.W. and Coburn, J.W., 2007. An acute bout of static stretching does not affect maximal eccentric isokinetic peak torque, the joint angle at peak torque, mean power,

electromyography, or mechanomyography. *Journal of Orthopaedic and Sports Physical Therapy*, 37(3), pp.130–139.

Cronin, J., Lawton, T., Harris, N., Kilding, A. and McMaster, D.T., 2017. A Brief Review of Handgrip Strength and Sport Performance. *Journal of Strength and Conditioning Research*, [online] 31(11), pp.3187–3217. Available at: <https://journals.lww.com/00124278-201711000-00029>.

Dailly, A.I., Sigrist, R., Kim, Y., Wolf, P., Erckens, H., Cerny, J., Luft, A., Gassert, R. and Sulzer, J., 2012. Can simple error sonification in combination with music help improve accuracy in upper limb movements? In: J.P. Desai, L.P.S. Jay and L.}. Zollo, eds. *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Proceedings of the IEEE RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics}. [online] IEEE.pp.1423–1427.

Daly, L.E. and Bourke, G.J., 2000. *Interpretation and Uses of Medical Statistics*. [online] Wiley. Available at: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470696750>.

Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., Kronland-Martinet, R., Ystad, S. and Velay, J.-L., 2015. The effect of real-time auditory feedback on learning new characters. *Human Movement Science*, [online] 43, pp.216–228.

Danna, J. and Velay, J.-L., 2017. On the Auditory-Proprioception Substitution Hypothesis: Movement Sonification in Two Deafferented Subjects Learning to Write New Characters. *Frontiers in Neuroscience*, [online] 11, p.137.

Dauty, M. and Rochcongar, P., 2001. Reproducibility of concentric and eccentric isokinetic strength of the knee flexors in elite volleyball players. *Isokinetics and Exercise Science*, 9, pp.129–132.

de Ruiter, C.J., Jones, D.A., Sargeant, A.J. and de Haan, A., 1999. Temperature Effect on the Rates of Isometric Force Development and Relaxation in the Fresh and Fatigued Human Adductor Pollicis Muscle. *Experimental Physiology*, [online] 84(6), pp.1137–1150. Available at: <http://doi.wiley.com/10.1111/j.1469-445X.1999.01895.x>.

de, Vet, Henrica C. W., et al. *Measurement in Medicine : A Practical Guide*, Cambridge University Press, 2011. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/anglia/detail.action?docID=802925.

Dempster, W.T., 1955. *Space Requirements of the Seated Operator: geometrical, kinematic, and mechanical aspects of the body, with special reference to the limbs. WADC Technical Report*. Wright-Patterson Air Force Base, Ohio.

Deones, V.L., Wiley, S.C. and Worrell, T., 1994. Assessment of Quadriceps Muscle Performance by a Hand-Held Dynamometer and an Isokinetic Dynamometer. *Journal of Orthopaedic & Sports Physical Therapy*, [online] 20(6), pp.296–301.

Dollings, H., Sandford, F., O'conaire, E. and Lewis, J.S., 2012. Shoulder Strength Testing: The Intra- and Inter-tester Reliability of Routine Clinical Tests, Using the PowertrackTM II Commander. *Shoulder and Elbow*, 4(2), pp.131–140.

Doss, W.S. and Karpovich, P. V., 1965. A comparison of concentric, eccentric, and isometric strength of elbow flexors. *Journal of Applied Physiology*, [online] 20(2), pp.351–353.

Dowman, L., McDonald, C.F., Hill, C.J., Lee, A., Barker, K., Boote, C., Glaspole, I., Goh, N., Southcott, A., Burge, A., Ndongo, R., Martin, A. and Holland, A.E., 2016. Reliability of the hand held dynamometer in measuring muscle strength in people with interstitial lung disease. *Physiotherapy*, [online] 102(3), pp.249–255.

Drouin, J.M., Valovich-McLeod, T.C., Shultz, S.J., Gansneder, B.M. and Perrin, D.H., 2004. Reliability and validity of the Biodex system 3 pro isokinetic dynamometer velocity, torque and position measurements. *European Journal of Applied Physiology*, 91(1), pp.22–29.

Dubus, G. and Bresin, R., 2013. A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE*, [online] 8(12), p.e82491.

Dyck, P.J., Boes, C.J., Mulder, D., Millikan, C., Windebank, A.J., Dyck, P.J.B. and Espinosa, R., 2005. History of standard scoring, notation, and summation of neuromuscular signs. A current survey and recommendation. *Journal of the Peripheral Nervous System*, 10(2), pp.158–173.

Dyer, J., Stapleton, P. and W. M. Rodger, M., 2015. Sonification as Concurrent Augmented Feedback for Motor Skill Learning and the Importance of Mapping Design. *The Open Psychology Journal*, [online] 8(1), pp.192–202.

Dyer, J., 2017. *Human Movement Sonification for Motor Skill Learning*. PhD. Queen's University Belfast.

Dyer, J.F., Stapleton, P. and Rodger, M., 2017a. Mapping sonification for perception and action in motor skill learning. *Frontiers in Neuroscience*, 11(AUG), pp.10–13.

Dyer, J.F., Stapleton, P. and Rodger, M.W.M., 2017b. Advantages of melodic over rhythmic movement sonification in bimanual motor skill learning. *Experimental Brain Research*, 235(10), pp.3129–3140.

Effenberg, A.O., 2005. Movement Sonification: Effects on Perception and Action. *IEEE Multimedia*, [online] 12(2), pp.53–59.

Effenberg, A.O., Fehse, U., Schmitz, G., Krueger, B. and Mechling, H., 2016. Movement Sonification: Effects on Motor Learning beyond Rhythmic Adjustments. *Frontiers in Neuroscience*, [online] 10(MAY).

El-Ashker, S., Carson, B.P., Ayala, F. and De Ste Croix, M., 2017. Sex-related differences in joint-angle-specific functional hamstring-to-quadriceps strength ratios. *Knee Surgery, Sports Traumatology, Arthroscopy*, 25(3), pp.949–957.

Emery, C.A., Owoeye, O.B.A., Räisänen, A.M., Befus, K., Hubkarao, T., Palacios-Derflingher, L. and Pasanen, K., 2022. The "SHRed Injuries Basketball" Neuromuscular Training Warm-up Program Reduces Ankle and Knee Injury Rates by 36% in Youth Basketball. *Journal of Orthopaedic & Sports Physical Therapy*, [online] 52(1), pp.40–48.

Engel, F.A., Faude, O., Kölling, S., Kellmann, M. and Donath, L., 2019. Verbal encouragement and between-day reliability during high-intensity functional strength and endurance performance testing. *Frontiers in Physiology*, 10(APR), pp.1–8.

Enoka, R.M., 1988. Muscle Strength and Its Development. *Sports Medicine*, [online] 6(3), pp.146–168.

Eriksson, M. and Bresin, R., 2010. Improving running mechanics by use of interactive sonification. In: *Proceedings of ISon*. [online] pp.95–98.

Eustace, S., Page, R.M. and Greig, M., 2019. Novel isokinetic dynamometry of the thigh musculature during anterior cruciate ligament rehabilitation in professional soccer: An explorative case study. *International Journal of Athletic Therapy and Training*, 24(2), pp.44–49.

Fan, E., Ciesla, N.D., Truong, A.D., Bhoopathi, V., Zeger, S.L. and Needham, D.M., 2010. Inter-rater reliability of manual muscle strength testing in ICU survivors and simulated patients. *Intensive Care Medicine*, 36(6), pp.1038–1043.

Fava, G.A., Tomba, E. and Sonino, N., 2012. Clinimetrics: The science of clinical measurements. *International Journal of Clinical Practice*, 66(1), pp.11–15.

Feaster, P., 2010. Edouard-Leon Scott De Martinville: An Annotated Discography. *ARSC Journal*, 41(1), pp.43–82.

Ferguson, J. and Brewster, S.A., 2018. Investigating Perceptual Congruence between Data and Display Dimensions in Sonification. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. [online] New York, NY, USA: ACM.pp.1–9.

Ferguson-Pell, M. and Cardi, M.D., 1993. Prototype Development and Comparative Evaluation of Wheelchair Pressure Mapping System. *Assistive Technology*, [online] 5(2), pp.78–91.

Fleiss, J.L., 1999. *The Design and Analysis of Clinical Experiments*. [online] Hoboken, NJ, USA: John Wiley & Sons, Inc. Available at: <http://doi.wiley.com/10.1002/9781118032923>.

Florence, J.M., Pandya, S., King, W.M., Robison, J.D., Baty, J., Miller, J.P., Schierbecker, J. and Signore, L.C., 1992. Intrarater reliability of manual muscle test (Medical Research Council scale) grades in Duchenne's muscular dystrophy. *Physical Therapy*, 72(2), pp.115–126.

Fujii, S., Lulic, T. and Chen, J.L., 2016. More Feedback Is Better than Less: Learning a Novel Upper Limb Joint Coordination Pattern with Augmented Auditory Feedback. *Frontiers in Neuroscience*, [online] 10(JUN), p.251.

Gallucci, J.G. and Challis, J.H., 2002. Examining the role of the gastrocnemius during the leg curl exercise. *Journal of Applied Biomechanics*, 18(1), pp.15–27.

Ganesan, S. and Anthony, L., 2012. Using the kinect to encourage older adults to exercise. In: *Proceedings of the 2012 ACM annual conference extended*

abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12. [online] New York, New York, USA: ACM Press.p.2297.

Garcia, M.A.C. and Souza, V.H., 2020. The (un)standardized use of handheld dynamometers on the evaluation of muscle force output. *Brazilian Journal of Physical Therapy*, [online] 24(1), pp.88–89.

Gardner, J., Webster, A. and Barry, J., 2018. Anticipating the clinical adoption of regenerative medicine: Building institutional readiness in the UK. *Regenerative Medicine*, 13(1), pp.29–39.

Garmer, K., Ylvén, J. and MariAnne Karlsson, 2004. User participation in requirements elicitation comparing focus group interviews and usability tests for eliciting usability requirements for medical equipment: a case study. *International Journal of Industrial Ergonomics*, [online] 33(2), pp.85–98.

Gary, R.A., Cress, M.E., Higgins, M.K., Smith, A.L. and Dunbar, S.B., 2011. Combined aerobic and resistance exercise program improves task performance in patients with heart failure. *Archives of Physical Medicine and Rehabilitation*, [online] 92(9), pp.1371–1381.

Gautrey, C.N., Watson, T. and Mitchell, A., 2013. The effect of isokinetic testing speed on the reliability of muscle fatigue indicators during a hip abductor-adductor fatigue protocol. *International Journal of Sports Medicine*, 34(7), pp.646–653.

Gentil, P., Del Vecchio, F.B., Paoli, A., Schoenfeld, B.J. and Bottaro, M., 2017. Isokinetic Dynamometry and 1RM Tests Produce Conflicting Results for Assessing Alterations in Muscle Strength. *Journal of Human Kinetics*, 56(1), pp.19–27.

Giansanti, D., Dozza, M., Chiari, L., Maccioni, G. and Cappello, A., 2009. Energetic assessment of trunk postural modifications induced by a wearable audio-biofeedback system. *Medical Engineering & Physics*, [online] 31(1), pp.48–54.

Giavarina, D., 2015. Understanding Bland Altman analysis. *Biochemia Medica*, 25(2), pp.141–151.

Gionfrida, L. and Roginska, A., 2017. A novel sonification approach to support the diagnosis of Alzheimer's dementia. *Frontiers in Neurology*, 8(DEC), pp.1–12.

Gleeson, N.P. and Mercer, T.H., 1996. The utility of isokinetic dynamometry in the assessment of human muscle function. *Sports Medicine*, 21(1), pp.18–34.

Gogte, K., Srivastav, P. and Miyaru, G.B., 2017. Effect of passive, active and combined warm up on lower limb muscle performance and dynamic stability in recreational sports players. *Journal of Clinical and Diagnostic Research*, 11(3), pp.YC05–YC08.

Goonetilleke, A., Modarres-Sadeghi, H. and Guiloff, R.J., 1994. Accuracy, reproducibility, and variability of hand-held dynamometry in motor neuron disease. *Journal of Neurology, Neurosurgery & Psychiatry*, [online] 57(3), pp.326–332.

Gorgas, A.-M., Schön, L., Dlapka, R., Doppler, J., Iber, M., Gradl, C., Kiselka, A., Siragy, T. and Horsak, B., 2017. Short-Term Effects of Real-Time Auditory Display (Sonification) on Gait Parameters in People with Parkinsons' Disease—A Pilot Study. In: J. Ibáñez, J. González-Vargas, J.M. Azorín, M. Akay and J.L. Pons, eds. *Converging Clinical and Engineering Research on Neurorehabilitation II*. [online] Cham: Springer International Publishing.pp.855–859. Available at: <http://link.springer.com/10.1007/978-3-319-46669-9_139>.

Greig, M. and Naylor, J., 2017. the Efficacy of Angle-Matched Isokinetic Knee Flexor and Extensor Strength Parameters in Predicting Agility Test Performance. *International Journal of Sports Physical Therapy*, 12(5), pp.728–736.

Großhauser, T., Bläsing, B., Spieth, C., Hermann, T., 2012. Wearable Sensor-Based Real-Time Sonification of Motion and Foot Pressure in Dance Teaching and Training. Journal of the Audio Engineering Society, 60(7/8), pp. 580-589.

Grooten, W.J.A. and Äng, B.O., 2010. Reliability of measurements of wrist extension force obtained with a Nicholas Manual Muscle Tester (NMMT). *Physiotherapy Theory and Practice*, 26(4), pp.281–287.

Guerra, J., Smith, L., Vicinanza, D., Stubbs, B., Veronese, N. and Williams, G., 2020. The use of sonification for physiotherapy in human movement tasks: A scoping review. *Science & Sports*, [online] 35(3), pp.119–129. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0765159720300046>.

Habets, B., Staal, J.B., Tijssen, M. and van Cingel, R., 2018. Intrarater reliability of the Humac NORM isokinetic dynamometer for strength measurements of the knee and shoulder muscles. *BMC Research Notes*, [online] 11(1), p.15. Available at: <https://doi.org/10.1186/s13104-018-3128-9>.

Haire, C.M., Tremblay, L., Vuong, V., Patterson, K.K., Chen, J.L., Burdette, J.H., Schaffert, N. and Thaut, M.H., 2021. Therapeutic Instrumental Music Training and Motor Imagery in Post-Stroke Upper-Extremity Rehabilitation: A Randomized-Controlled Pilot Study. *Archives of Rehabilitation Research and Clinical Translation*, [online] 3(4), p.100162.

Hansen, E.M., McCartney, C.N., Sweeney, R.S., Palimenio, M.R. and Grindstaff, T.L., 2015. Hand-held Dynamometer Positioning Impacts Discomfort During Quadriceps Strength Testing: A Validity and Reliability Study. *International journal of sports physical therapy*, [online] 10(1), pp.62–8.

Hayes, K.W. and Falconer, J., 1992. Reliability of hand-held dynamometry and its relationship with manual muscle testing in patients with osteoarthritis in the knee. *Journal of Orthopaedic and Sports Physical Therapy*, 16(3), pp.145–149.

HCPC, 2013. *Standards of proficiency - Physiotherapists*. [online] Health and Care Professions Council. Available at: <http://www.hpc-uk.org/globalassets/resources/standards/standards-of-proficiency---physiotherapists.pdf?v=637106257580000000>.

Hébert, L.J., Remec, J.-F., Saulnier, J., Vial, C. and Puymirat, J., 2010. The use of muscle strength assessed with handheld dynamometers as a non-invasive biological marker in myotonic dystrophy type 1 patients: a multicenter study. *BMC Musculoskeletal Disorders*, [online] 11(1), p.72.

Hegg, J.C., Middleton, J., Robertson, B.L. and Kennedy, B.P., 2018. The sound of migration: exploring data sonification as a means of interpreting multivariate salmon movement datasets. *Heliyon*, [online] 4(2), p.e00532.

Hendy, A.M., Spittle, M. and Kidgell, D.J., 2012. Cross education and immobilisation: Mechanisms and implications for injury rehabilitation. *Journal of Science and Medicine in Sport*, [online] 15(2), pp.94–101. Available at: <http://dx.doi.org/10.1016/j.jsams.2011.07.007>.

Hermann, T., 2008. Taxonomy and Definitions for Sonification and Auditory Display. In: *Proceedings of the 14th International Conference on Auditory Display, Paris, France.*

Hermann, T., Hunt, A. and Neuhoff, J.G., 2011. *The Sonification Handbook*. 1st ed. [online] Logos Publishing House, Berlin. Available at: <https://sonification.de/handbook/>.

Hermann, T., 2011. Model-Based Sonification. In: T. Hermann, A. Hunt and J. Neuhoff, ed. 2011. *The sonification handbook*. Logos Verlag, Berlin, Germany. Ch. 16.

Hill, T.R., Gjellesvik, T.I., Moen, P.M.R., Tørhaug, T., Fimland, M.S., Helgerud, J. and Hoff, J., 2012. Maximal strength training enhances strength and functional performance in chronic stroke survivors. *American Journal of Physical Medicine and Rehabilitation*, 91(5), pp.393–400.

Hill, T., 2014. *Torque it up: reliability and variability of isokinetic peak torque and angle of peak torque*. University of Otago.

Hirschmann, M.T., Wind, B., Amsler, F. and Gross, T., 2010. Reliability of shoulder abduction strength measure for the constant-murley score. *Clinical Orthopaedics and Related Research*, 468(6), pp.1565–1571.

Hislop, H. J., and Montgomery, J., 2007. *Daniels and Worthingham's muscle testing: techniques of manual examination*. 8th ed.,St. Louis, Mo.: Saunders / Elsevier

Holt, K.L., Raper, D.P., Boettcher, C.E., Waddington, G.S. and Drew, M.K., 2016. Hand-held dynamometry strength measures for internal and external rotation demonstrate superior reliability, lower minimal detectable change and higher correlation to isokinetic dynamometry than externally-fixed dynamometry of the shoulder. *Physical Therapy in Sport*, [online] 21, pp.75–81. Available at: <http://dx.doi.org/10.1016/j.ptsp.2016.07.001>.

Hopkins, K. D. and Stanley, J. C.,(1981). *Educational and psychological measurement and evaluation (6th ed.)*. Englewood Cliffs, NJ: Prentice Hall.

Hopkins, W.G., Marshall, S.W., Batterham, A.M. and Hanin, J., 2009. Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, 41(1), pp.3–12.

Huber, D.M., 2012. *The MIDI Manual*. [online] Routledge. Available at: <https://www.taylorfrancis.com/books/9781136119187>.

Hughes, L., Rosenblatt, B., Haddad, F., Gissane, C., McCarthy, D., Clarke, T., Ferris, G., Dawes, J., Paton, B. and Patterson, S.D., 2019. Comparing the Effectiveness of Blood Flow Restriction and Traditional Heavy Load Resistance Training in the Post-Surgery Rehabilitation of Anterior Cruciate Ligament Reconstruction Patients: A UK National Health Service Randomised Controlled Trial. *Sports Medicine*, [online] 49(11), pp.1787–1805.

Hunt, A., Hermann, T. and Pauletto, S., 2004. Interacting with Sonification Systems: Closing the Loop. In: *Proceedings of the Eighth International Conference on Information Visualisation (IV'04)*. [online] IEEE. pp.879–884.

Imada, A.S., 1991. *The rationale and tools of participatory ergonomics*. In: Noro, K., Imada, A. (Eds.), Participatory Ergonomics. Taylor and Francis, London, pp. 30–49

International Organization for Standardization, ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, 2018

Ishøi, L., Sørensen, C.N., Kaae, N.M., Jørgensen, L.B., Hölmich, P. and Serner, A., 2016. Large eccentric strength increase using the Copenhagen Adduction exercise in football: A randomized controlled trial. *Scandinavian Journal of Medicine and Science in Sports*, 26(11), pp.1334–1342.

Jackson, S.M., Cheng, M.S., Smith, A.R. and Kolber, M.J., 2017. Intrarater reliability of hand held dynamometry in measuring lower extremity isometric strength using a portable stabilization device. *Musculoskeletal Science and Practice*, [online] 27, pp.137–141.

Jaeschke, R., Singer, J. and Guyatt, G.H., 1989. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, 10(4), pp.407–415.

Jakus, G., Stojmenova, K., Tomažič, S. and Sodnik, J., 2017. A system for efficient motor learning using multimodal augmented feedback. *Multimedia Tools and Applications*, [online] 76(20), pp.20409–20421.

Janssen, J.C. and Le-Ngoc, L., 2009. Intratester Reliability and Validity of Concentric Measurements Using a New Hand-Held Dynamometer. *Archives of Physical Medicine and Rehabilitation*, [online] 90(9), pp.1541–1547.

Jesus, T.S., Gianola, S., Castellini, G., Colquhoun, H. and Brooks, D., 2020. Evolving trends in physiotherapy research publications between 1995 and 2015. *Physiotherapy Canada*, 72(2), pp.122–131.

Kairy, D., Tousignant, M., Leclerc, N., Côté, A.-M., Levasseur, M. and Researchers, the, 2013. The Patient's Perspective of in-Home Telerehabilitation Physiotherapy Services Following Total Knee Arthroplasty. *International Journal of Environmental Research and Public Health*, [online] 10(9), pp.3998–4011.

Kannus, P., 1990. Relationships between peak torque, peak angular impulse, and average power in the thigh muscles of subjects with knee damage. *Research Quarterly for Exercise and Sport*, 61(2), pp.141–145.

Kannus, P., 1994. Isokinetic Evaluation of Muscular Performance. *International Journal of Sports Medicine*, 15(S 1), pp.S11–S18.

Karabay, D., Yesilyaprak, S.S. and Sahiner Picak, G., 2020. Reliability and validity of eccentric strength measurement of the shoulder abductor muscles using a hand-held dynamometer. *Physical Therapy in Sport*, [online] 43, pp.52–57.

Keep, H., Luu, L., Berson, A. and Garland, S.J., 2016. Validity of the handheld dynamometer compared with an isokinetic dynamometer in measuring peak hip extension strength. *Physiotherapy Canada*, 68(1), pp.15–22.

Kellis, E. and Baltzopoulos, V., 1996. Gravitational moment correction in isokinetic dynamometry using anthropometric data. *Medicine and Science in Sports and Exercise*, 28(7), pp.900–907.

Kelln, B.M., McKeon, P.O., Gontkof, L.M. and Hertel, J., 2008. Hand-held dynamometry: Reliability of lower extremity muscle testing in healthy, physically active, young adults. *Journal of Sport Rehabilitation*, 17(2), pp.160–170.

Kamberelis, G., and Dimitriadis, G., 2005. Focus groups: Strategic articulations of pedagogy, politics, and inquiry. In: N. K. Denzin, & Y. S. Lincoln, Eds. 2005.

The Sage Handbook of Qualitative Research, 3rd ed.. Thousand Oaks, CA: Sage Publications Inc. pp.887–907.

Kang, H., 2013. The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), pp.402–406.

Kendall, F., McCreary, E., Provance, P., Rodgers, M. and Romani, W., 2005. *Muscles: Testing and Function With Posture and Pain*. 5th ed. Philadelphia: Lippincott Williams & Wilkins.

Kenyon, G.P. and Thaut, M.H. 2003. Rhythmic-driven optimization of motor control. In: Thaut, M.H. (ed.) *Rhythm, music and the brain: scientific foundations and clinical applications (studies on new music research)*. New York and London: Routledge Chapman & Hall, 2008, pp.85–112.

Kilmer, D.D., McCrory, M.A., Wright, N.C., Rosko, R.A., Kim, H.R. and Aitkens, S.G., 1997. Hand-held dynamometry reliability in persons with neuropathic weakness. *Archives of Physical Medicine and Rehabilitation*, 78(12), pp.1364–1368.

Kim, W.K., Kim, D.K., Seo, K.M. and Kang, S.H., 2014. Reliability and validity of isometric knee extensor strength test with hand-held dynamometer depending on its fixation: A pilot study. *Annals of Rehabilitation Medicine*, 38(1), pp.84–93.

Knapik, J.J., Wright, J.E., Mawdsley, R.H. and Braun, J., 1983. Isometric, isotonic, and isokinetic torque variations in four muscle groups through a range of joint motion. *Physical Therapy*, 63(6), pp.938–947.

Knepler, C. and Bohannon, R.W., 1998. Subjectivity of Forces Associated with Manual-Muscle Test Grades of 3+, 4-, and 4. *Perceptual and Motor Skills*, [online] 87(3_suppl), pp.1123–1128.

Knippenberg, E., Verbrugghe, J., Lamers, I., Palmaers, S., Timmermans, A. and Spooren, A., 2017. Markerless motion capture systems as training device in neurological rehabilitation: a systematic review of their use, application, target population and efficacy. *Journal of NeuroEngineering and Rehabilitation*, [online] 14(1), p.61.

Koo, T.K. and Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), pp.155–163.

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M. and Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), pp.96–106.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, R., Bargar, J., Barrass, S., Berger, J., Evreinov, G., Tecumseh Fitch, W., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M. and Tipei, S., 1998. *Sonification Report: Status of the Field and Research Agenda*. [online] Available at: <http://sonify.psych.gatech.edu/publications/pdfs/1999-NSF-Report.pdf>.

Kreighbaum, E. and Barthels K.M., 1996. *Biomechanics: A Qualitative Approach for Studying Human Movement*. 4th ed. Needham Heights, MA: Allyn & Bacon.

Królikowska, A., Reichert, P., Czamara, A. and Krzemińska, K., 2019. Peak torque angle of anterior cruciate ligament-reconstructed knee flexor muscles in patients with semitendinosus and gracilis autograft is shifted towards extension regardless of the postoperative duration of supervised physiotherapy. *PLoS ONE*, 14(2), pp.1–22.

Kronborg, L., Bandholm, T., Palm, H., Kehlet, H. and Kristensen, M.T., 2017. Effectiveness of acute in-hospital physiotherapy with knee-extension strength training in reducing strength deficits in patients with a hip fracture: A randomised controlled trial. *PLoS ONE*, 12(6), pp.1–18.

Lam, A.W.K., Varona-Marin, D., Li, Y., Fergenbaum, M. and Kulic, D., 2016. Automated Rehabilitation System: Movement Measurement and Feedback for Patients and Physiotherapists in the Rehabilitation Clinic. *Human-Computer Interaction*, [online] 31(3–4), pp.294–334.

Lauszuz, Kristian, 2020. Kalman Filter library. *GitHub*, Available at: https://github.com/TKJElectronics/KalmanFilter.

Lee, T.H., Park, K.S., Lee, D.G. and Lee, N.G., 2012. Concurrent validity by comparing EMG activity between manual muscle testing, handheld dynamometer, and stationary dynamometer in testing of maximal isometric quadriceps contraction. *Journal of Physical Therapy Science*, 24(12), pp.1219–1223.

Legg, H.S., Spindor, J., Dziendzielowski, R., Sharkey, S., Lanovaz, J.L., Farthing, J.P. and Arnold, C.M., 2020. The reliability and validity of novel clinical strength measures of the upper body in older adults. *Hand Therapy*, 25(4), pp.130–138.

Leggin, B.G., Neuman, R.M., Iannotti, J.P., Williams, G.R. and Thompson, E.C., 1996. Intrarater and interrater reliability of three isometric dynamometers in assessing shoulder strength. *Journal of Shoulder and Elbow Surgery*, [online] 5(1), pp.18–24.

Le-Ngoc, L. and Janssen, J., 2012. Validity and reliability of a hand-held dynamometer for dynamic muscle strength assessment. In: Chong-Tae Kim, ed. *Rehabilitation Medicine*. [online] InTech. Available at: <http://www.intechopen.com/books/rehabilitation-medicine/validity-and-reliability-of-a-hand-held-dynamometer-for-dynamic-muscle-strength-assessment>.

Lesnak, J., Anderson, D., Farmer, B., Katsavelis, D. and Grindstaff, T.L., 2019. Validity of hand-held dynamometry in measuring quadriceps strength and rate of torque development. *International Journal of Sports Physical Therapy*, [online] 14(2), pp.180–187.

Lexell, J.E. and Downham, D.Y., 2005. How to Assess the Reliability of Measurements in Rehabilitation. *American Journal of Physical Medicine & Rehabilitation*, [online] 84(9), pp.719–723.

Li, R.C., Jasiewicz, J.M., Middleton, J., Condie, P., Barriskill, A., Hebnes, H. and Purcell, B., 2006. The Development, Validity, and Reliability of a Manual Muscle Testing Device With Integrated Limb Position Sensors. *Archives of Physical Medicine and Rehabilitation*, [online] 87(3), pp.411–417. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S0003999305014231>.

Lieber, R.L. and Ward, S.R., 2011. Skeletal muscle design to meet functional demands. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1570), pp.1466–1476.

Ludbrook, J., 2002. Statistical Techniques For Comparing Measurers And Methods Of Measurement: A Critical Review. *Clinical and Experimental Pharmacology and Physiology*, [online] 29(7), pp.527–536.

Ludbrook, J., 2010. Confidence in Altman-Bland plots: A critical review of the method of differences. *Clinical and Experimental Pharmacology and Physiology*, 37(2), pp.143–149.

Ludbrook, J., 2012. A primer for biomedical scientists on how to execute Model II linear regression analysis. *Clinical and Experimental Pharmacology and Physiology*, 39(4), pp.329–335.

Ludovico, L.A. and Presti, G., 2016. The sonification space: A reference system for sonification tasks. *International Journal of Human Computer Studies*, [online] 85, pp.72–77.

Lynall, R.C., Zukowski, L.A., Plummer, P. and Mihalik, J.P., 2017. Reliability and validity of the protokinetics movement analysis software in measuring center of pressure during walking. *Gait & Posture*, [online] 52, pp.308–311.

Maes, P.J., Lorenzoni, V. and Six, J., 2018. The SoundBike: musical sonification strategies to enhance cyclists' spontaneous synchronization to external music. *Journal on Multimodal User Interfaces*, [online] 13(3), pp.155–166.

Mafi, P., Mafi, R., Hindocha, S., Griffin, M. and Khan, W., 2012. A Systematic Review of Dynamometry and its Role in Hand Trauma Assessment. *The Open Orthopaedics Journal*, [online] 6(1), pp.95–102.

Maffiuletti, N.A., Bizzini, M., Desbrosses, K., Babault, N. and Munzinger, U., 2007. Reliability of knee extension and flexion measurements using the Con-Trex isokinetic dynamometer. *Clinical Physiology and Functional Imaging*, [online] 27(6), pp.346–353.

Mahony, K., Hunt, A., Daley, D., Sims, S. and Adams, R., 2009. Inter-tester reliability and precision of manual muscle testing and hand-held dynamometry in lower limb muscles of children with spina bifida. *Physical and Occupational Therapy in Pediatrics*, 29(1), pp.44–59.

Martien, S., Delecluse, C., Boen, F., Seghers, J., Pelssers, J., Van Hoecke, A.S. and Van Roie, E., 2015. Is knee extension strength a better predictor of functional performance than handgrip strength among older adults in three different settings? *Archives of Gerontology and Geriatrics*, [online] 60(2), pp.252–258. Available at: <http://dx.doi.org/10.1016/j.archger.2014.11.010>.

Martin, H.J., Yule, V., Syddall, H.E., Dennison, E.M., Cooper, C. and Aihie Sayer, A., 2006. Is hand-held dynamometry useful for the measurement of quadriceps strength in older people? A comparison with the gold standard biodex dynamometry. *Gerontology*, 52(3), pp.154–159.

Martin, R.L., Cibulka, M.T., Bolgla, L.A., Koc, T.A., Loudon, J.K., Manske, R.C., Weiss, L., Christoforetti, J.J. and Heiderscheit, B.C., 2022. Hamstring Strain Injury in Athletes. *Journal of Orthopaedic & Sports Physical Therapy*, [online] 52(3), pp.CPG1–CPG44.

Martín-San Agustín, R., Benítez-Martínez, J.C., Castillo-Ballesta, L., Gacto-Sánchez, M. and Medina-Mirapeix, F., 2020. Validity, Reliability, and Sensitivity to Change of DiCI for the Strength Measurement of Knee and Hip Muscles. *Measurement in Physical Education and Exercise Science*, [online] 24(4), pp.303–311.

Martins, J., Da Silva, J.R., Da Silva, M.R.B. and Bevilaqua-Grossi, D., 2017. Reliability and validity of the belt-stabilized handheld dynamometer in hip-and knee-strength tests. *Journal of Athletic Training*, 52(9), pp.809–819.

Matsubara, M., Iguchi, M., Oba, T., Kadone, H., Terasawa, H. and Suzuki, K., 2015. An Instrumented Ankle-Foot Orthosis with Auditory Biofeedback for Blind and Sighted Individuals. *IEEE MultiMedia*, [online] PP(99), pp.1–1.

Mayer, F., Horstmann, T., Kranenberg, U., Röcker, K. and Dickhuth, H.-H., 1994. Reproducibility of Isokinetic Peak Torque and Angle at Peak Torque in the Shoulder Joint. *International Journal of Sports Medicine*, [online] 15(S 1), pp.S26–S31.

McBride, J., 2016. Biomechanics of Resistance Exercise. In Haff, G. & Triplett, N. T. 2016. *Essentials of strength training and conditioning*. Fourth Edition. Champaing, IL: Human Kinetics.

McDonnell, B., Stillwell, S., Hart, S. and Davis, R.B., 2018. Breaking Down Barriers to the Utilization of Standardized Tests and Outcome Measures in Acute Care Physical Therapist Practice: An Observational Longitudinal Study. *Physical Therapy*, [online] 98(6), pp.528–538.

Mendonca, L.D., Bittencourt, N.F.N., Barreto, R.A., Paiva, T.F., Porto, R.F., Silva, A.A. and Fonseca, S.T., 2011. Correlation between isokinetic profile and knee

injuries in male volleyball athletes. *British Journal of Sports Medicine*, [online] 45(4), pp.345–345.

Mentiplay, B.F., Perraton, L.G., Bower, K.J., Adair, B., Pua, Y.H., Williams, G.P., McGaw, R. and Clark, R.A., 2015. Assessment of lower limb muscle strength and power using hand-held and fixed dynamometry: A reliability and validity study. *PLoS ONE*, 10(10), pp.1–18.

Mezzarobba, S., Grassi, M., Pellegrini, L., Catalan, M., Krüger, B., Stragapede, L., Manganotti, P. and Bernardis, P., 2020. Action observation improves sit-to-walk in patients with Parkinson's disease and freezing of gait. Biomechanical analysis of performance. *Parkinsonism & Related Disorders*, [online] 80, pp.133–137.

Micheletti, J.K., Andersen, L.L., Machado, A.F., Lopes, J.S.S., Souto, L.R. and Pastre, C.M., 2020. Reliability of shoulder abduction muscle endurance measured using isokinetic dynamometry. *Science & Sports*.

Miyaguchi, K. and Demura, S., 2008. Relationships between Stretch-shortening Cycle Performance and Maximum Muscle Strength. *Journal of Strength and Conditioning Research*, [online] 22(1), pp.19–24. Available at: <https://journals.lww.com/00124278-200801000-00005>.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M. and de Vet, H.C.W., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, [online] 63(7), pp.737–745.

Montgomery, L.C., Douglass, L.W. and Deuster, P.A., 1989. Reliability of an isokinetic test of muscle strength and endurance. *Journal of Orthopaedic and Sports Physical Therapy*, 10(8), pp.315–322.

Morrissey, M.C., 1987. The relationship between peak torque and work of the quadriceps and hamstrings after meniscectomy. *Journal of Orthopaedic and Sports Physical Therapy*, 8(8), pp.405–408.

Mouelhi, Y., Jouve, E., Castelli, C. and Gentile, S., 2020. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and Quality of Life Outcomes*, 18(1), pp.1–17.

Muff, G., Dufour, S., Meyer, A., Severac, F., Favret, F., Geny, B., Lecocq, J. and Isner-Horobeti, M.E., 2016. Comparative assessment of knee extensor and flexor muscle strength measured using a hand-held vs. Isokinetic dynamometer. *Journal of Physical Therapy Science*, 28(9), pp.2445–2451.

Mulroy, S. and Lassen, D., 1997. The Ability of Male and Female Clinicians to effectively Test Knee Extension Strength Using Manual Muscle Testing. *Journal of Orthopaedic & Sports Physical Therapy*, 26(4), pp.192–199.

Munn, Z., Stern, C., Aromataris, E., Lockwood, C. and Jordan, Z., 2018. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*, [online] 18(1), p.5. Available at: <https://doi.org/10.1186/s12874-017-0468-4>.

Nagatomi, T., Mae, T., Nagafuchi, T., Yamada, S., Nagai, K. and Yoneda, M., 2017. Shoulder manual muscle resistance test cannot fully detect muscle weakness. *Knee Surgery, Sports Traumatology, Arthroscopy*, 25(7), pp.2081–2088.

Necula, Bogdan, 2021. HX711. *GitHub*, Available at:https://github.com/bogde/HX711/.

Newbold, J.W., Bianchi-Berthouze, N., Gold, N.E., Tajadura-Jiménez, A. and Williams, A.C., 2016. Musically Informed Sonification for Chronic Pain Rehabilitation. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. [online] New York, NY, USA: ACM.pp.5698–5703.

Newbold, J.W., Bianchi-Berthouze, N. and Gold, N.E., 2017. Musical Expectancy in Squat Sonification for People Who Struggle with Physical Activity. In: *Proceedings of the 23rd International Conference on Auditory Display - ICAD 2017*. [online] Arlington, Virginia: The International Community for Auditory Display.pp.65–72. Available at: <http://hdl.handle.net/1853/58350>.

Nielsen, J., 1994. Usability Engineering. [online] Computing Handbook, Third Edition. San Francisco, CA, United States: Morgan Kaufmann Publishers Inc.340. Available at: <http://www.crcnetbase.com/doi/abs/10.1201/b16768-38>.

Nikmaram, N., Scholz, D.S., Großbach, M., Schmidt, S.B., Spogis, J., Belardinelli, P., Müller-Dahlhaus, F., Remy, J., Ziemann, U., Rollnik, J.D. and Altenmüller, E., 2019. Musical Sonification of Arm Movements in Stroke Rehabilitation Yields Limited Benefits. *Frontiers in Neuroscience*, 13(December), pp.1–14.

Nordin, F., Nyberg, A. and Sandberg, C., 2020. Concurrent validity of a fixated hand-held dynamometer for measuring isometric knee extension strength in adults with congenital heart disease. *European Journal of Physiotherapy*, 22(4), pp.206–211.

Noreau, L. and Vachon, J., 1998. Comparison of three methods to assess muscular strength in individuals with spinal cord injury. *Spinal Cord*, [online] 36(10), pp.716–723.

Nugent, E.P., Snodgrass, S.J. and Callister, R., 2015. The effect of velocity and familiarisation on the reproducibility of isokinetic dynamometry. *Isokinetics and Exercise Science*, 23(3), pp.205–214.

O´Brien, B., Vidal, A., Bringoux, L. and Bourdin, C., 2020. Developing Movement Sonification for Sports Performance. In: I.S.W. (ISON 2019), ed. *Proceedings of ISon 2019, 6thInteractive Sonification Work*. KTH Royal Institute of Technology.pp.124–129.

O'Brien, B., Juhas, B., Bieńkiewicz, M., Buloup, F., Bringoux, L. and Bourdin, C., 2021. Sonification of Golf Putting Gesture Reduces Swing Movement Variability in Novices. *Research Quarterly for Exercise and Sport*, [online] 92(3), pp.301–310.

Oh, J.S., Kang, M.H. and Dvir, Z., 2016. Reproducibility of isometric shoulder protraction and retraction strength measurements in normal subjects and individuals with winged scapula. *Journal of Shoulder and Elbow Surgery*, [online] 25(11), pp.1816–1823.

O'Neill, S., Barry, S. and Watson, P., 2019. Plantarflexor strength and endurance deficits associated with mid-portion Achilles tendinopathy: The role of soleus. *Physical Therapy in Sport*, [online] 37, pp.69–76.

Orri, J.C. and Darden, G.F., 2008. Reliability and validity of the iSAM 9000 isokonetic dynamometer. *Journal of Strength and Conditioning Research*, 22(1), pp.310–317.

Palmer, M.L. and Epler, M.E., 1998. Fundamentals of Musculoskeletal Assessment Techniques (2nd Edition). Lippincott-Raven Publishers, Philadelphia.

Parcell, A.C., Sawyer, R.D., Tricoli, V.A. and Chinevere, T.D., 2002. Minimum rest period for strength recovery. *Medicine & Science in Sport & Exercise*, 34(6), pp.1018–1022.

Parmentier, F.B.R., 2014. The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, 78(3), pp.321–338.

Pauletto, S. and Hunt, A., 2009. Interactive sonification of complex data. *International Journal of Human-Computer Studies*, [online] 67(11), pp.923–933.

Pearn, J., 1978. Two early dynamometers. *Journal of the Neurological Sciences*, [online] 37(1–2), pp.127–134.

Peek, K., 2014. Muscle strength in adults with spinal cord injury: a systematic review of manual muscle testing, isokinetic and hand held dynamometry clinimetrics. *JBI Database of Systematic Reviews and Implementation Reports*, 12(5), pp.349–429.

Peres, S.C., Verona, D., Nisar, T. and Ritchey, P., 2017. Towards a systematic approach to real-time sonification design for surface electromyography. *Displays*, [online] 47, pp.25–31.

Perrin, D.H., 1986. Reliability of isokinetic measures. *Athletic training*, 21, pp.319–321.

Perrin, D.H., Robertson, R.J. and Ray, R.L., 1987. Bilateral isokinetic peak torque, torque acceleration energy, power, and work relationships in athletes and nonathletes. *Journal of Orthopaedic and Sports Physical Therapy*, 9(5), pp.184–189.

Perry, J., Weiss, W.B., Burnfield, J.M. and Gronley, J.K., 2004. The supine hip extensor manual muscle test: A reliability and validity study. *Archives of Physical Medicine and Rehabilitation*, 85(8), pp.1345–1350.

Pescatello, L.S., Arena, R., Riebe, D. and Thompson, P.D., 2014. *ACSM's Guidelines for Exercise Testing and Prescription*. 9th ed. hiladelphia : Wolters Kluwer/Lippincott Williams & Wilkins Health.

Petty, N.J., 2011. *Neuromusculoskeletal Examination and Assessment*. 4th ed. Churchill Livingstone.

Pfister, A., West, A.M., Bronner, S. and Noah, J.A., 2014. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of Medical Engineering & Technology*, [online] 38(5), pp.274–280.

Pfister, P.B., De Bruin, E.D., Sterkele, I., Maurer, B., De Bie, R.A. and Knols, R.H., 2018. Manual muscle testing and hand-held dynamometry in people with inflammatory myopathy: An intra- and interrater reliability and validity study. *PLoS ONE*, 13(3), pp.1–22.

Pham, M.T., Rajić, A., Greig, J.D., Sargeant, J.M., Papadopoulos, A. and McEwen, S.A., 2014. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research Synthesis Methods*, [online] 5(4), pp.371–385.

Pieters, D., Witvrouw, E., Wezenbeek, E. and Schuermans, J., 2020. Value of isokinetic strength testing for hamstring injury risk assessment: Should the 'strongest' mates stay ashore? *European Journal of Sport Science*, [online] pp.1–12.

Poon, E.G., Jha, A.K., Christino, M., Honour, M.M., Fernandopulle, R., Middleton, B., Newhouse, J., Leape, L., Bates, D.W., Blumenthal, D. and Kaushal, R., 2006. Assessing the level of healthcare information technology adoption in the United States: A snapshot. *BMC Medical Informatics and Decision Making*, 6, pp.1–9.

Portney, L.G. and Watkins, M.P., 2015. *Foundation Clinical Research: Applications to practice*. 3rd ed. Upper Saddle River, N.J: Pearson/Prentice Hall.

Rausch-Osthoff, A.K., Kohler, M., Sievi, N.A., Clarenbach, C.F. and Van Gestel, A.J.R., 2014. Association between peripheral muscle strength, exercise performance, and physical activity in daily life in patients with Chronic Obstructive Pulmonary Disease. *Multidisciplinary Respiratory Medicine*, 9(1), pp.1–7.

Reh, J., Hwang, T.H., Schmitz, G. and Effenberg, A.O., 2019. Dual mode gait sonification for rehabilitation after unilateral hip arthroplasty. *Brain Sciences*, 9(3).

Rendos, N.K., Harriell, K., Qazi, S., Regis, R.C., Alipio, T.C. and Signorile, J.F., 2019. Variations in Verbal Encouragement Modify Isokinetic Performance. *Journal of Strength and Conditioning Research*, [online] 33(3), pp.708–716.

Ribeiro, F., Lépine, P.A., Garceau-Bolduc, C., Coats, V., Allard, É., Maltais, F. and Saey, D., 2015. Test-retest reliability of lower limb isokinetic endurance in COPD: A comparison of angular velocities. *International Journal of COPD*, 10, pp.1163–1172.

Romero-Franco, N., Jiménez-Reyes, P. and Montaño-Munuera, J.A., 2017. Validity and reliability of a low-cost digital dynamometer for measuring isometric strength of lower limb. *Journal of Sports Sciences*, [online] 35(22), pp.2179–2184.

Romero-Franco, N., Fernández-Domínguez, J.C., Montaño-Munuera, J.A., Romero-Franco, J. and Jiménez-Reyes, P., 2019. Validity and reliability of a low-cost dynamometer to assess maximal isometric strength of upper limb: Low cost dynamometry and isometric strength of upper limb. *Journal of Sports Sciences*, [online] 37(15), pp.1787–1793.

Ross, J., Stevenson, F., Lau, R. and Murray, E., 2016. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implementation Science*, [online] 11(1), p.146.

Ross, J., Stevenson, F., Dack, C., Pal, K., May, C., Michie, S., Barnard, M. and Murray, E., 2018. Developing an implementation strategy for a digital health intervention: An example in routine healthcare 11 Medical and Health Sciences 1117 Public Health and Health Services. *BMC Health Services Research*, 18(1), pp.1–13.

Rothgangel, A., Kanera, I., van den Heuvel, R., Wientgen, M., Jamin, G., Lenssen, T. and Braun, S., 2020. Physiotherapists' clinical use and acceptance of a telemonitoring platform during anterior cruciate ligament rehabilitation: a prospective clinical study. *Disability and Rehabilitation: Assistive Technology*, [online] pp.1–8.

Rushton, A., Heneghan, N.R., Heap, A., White, L., Calvert, M. and Goodwin, P.C., 2017. Patient and physiotherapist perceptions of rehabilitation following primary lumbar discectomy: A qualitative focus group study embedded within an external pilot and feasibility trial. *BMJ Open*, 7(4), pp.1–11.

Russell, T.G., Buttrum, P., Wootton, R. and Jull, G.A., 2011. Internet-Based Outpatient Telerehabilitation for Patients Following Total Knee Arthroplasty. *The Journal of Bone and Joint Surgery-American Volume*, [online] 93(2), pp.113–120.

Sale, D.G., 1991. Testing strength and power (pp. 21–75). In: MacDougal JD, Wenger HA, Green HJ, editors. *Physiological testing of the high performance athlete*. Champaign, IL: Human Kinetics.

Samosawala, N.R., Vaishali, K. and Kalyana, B.C., 2016. Measurement of muscle strength with handheld dynamometer in Intensive Care Unit. *Indian Journal of Critical Care Medicine*, [online] 20(1), pp.21–26.

Sargeant, A.J., 2007. Structural and functional determinants of human muscle power. *Experimental Physiology*, 92(2), pp.323–331.

Scarf, F.L., Gurnett, D.A., Kurth, W.S. and Poynter, R.L., 1982. Voyager 2 Plasma Wave Observations at Saturn. *Science*, [online] 215(4532), pp.587–594. Available at: <https://www.sciencemag.org/lookup/doi/10.1126/science.215.4532.587>.

Schaffert, N., Mattes, K. and Effenberg, A.O., 2011. An investigation of online acoustic information for elite rowers in on-water training conditions. *Journal of Human Sport and Exercise*, [online] 6(2), pp.392–405.

Schaffert, N. and Mattes, K., 2016. Influence of acoustic feedback on boat speed and crew synchronization in elite junior rowing. *International Journal of Sports Science & Coaching*, [online] 11(6), pp.832–845.

Schaffert, N., Godbout, A., Schlueter, S. and Mattes, K., 2017. Towards an application of interactive sonification for the forces applied on the pedals during cycling on the Wattbike ergometer. *Displays*, [online] 50, pp.41–48.

Schaffert, N., Janzen, T.B., Mattes, K. and Thaut, M.H., 2019. A review on the relationship between sound and movement in sports and rehabilitation. *Frontiers in Psychology*, 10(FEB), pp.1–20.

Schmitz, G., Mohammadi, B., Hammer, A., Heldmann, M., Samii, A., Münte, T.F. and Effenberg, A.O., 2013. Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neuroscience*, [online] 14(1), p.32.

Schmitz, G., Kroeger, D. and Effenberg, A.O., 2014. A mobile sonification system for stroke rehabilitation. In: *The 20th International Conference on Auditory Display (ICAD-2014)*. New York, NY: Georgia Institute of Technology.

Scholz, D.S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J.D., Großbach, M. and Altenmller, E.O., 2014. Sonification as a possible stroke rehabilitation strategy. *Frontiers in Neuroscience*, [online] 8(OCT), p.332.

Scholz, D.S., Rhode, S., Großbach, M., Rollnik, J. and Altenmüller, E., 2015. Moving with music for stroke rehabilitation: a sonification feasibility study. *Annals of the New York Academy of Sciences*, [online] 1337(1), pp.69–76.

Scholz, D.S., Rohde, S., Nikmaram, N., Brückner, H.-P., Großbach, M., Rollnik, J.D. and Altenmüller, E.O., 2016. Sonification of Arm Movements in Stroke Rehabilitation – A Novel Approach in Neurologic Music Therapy. *Frontiers in Neurology*, [online] 7(JUN), p.106.

Schrama, P.P.M., Stenneberg, M.S., Lucas, C. and Van Trijffel, E., 2014. Intraexaminer reliability of hand-held dynamometry in the upper extremity: A systematic review. *Archives of Physical Medicine and Rehabilitation*, [online] 95(12), pp.2444–2469.

Schuyler, N. and Eun-Jin, K., 2016. Structures in sound: Analysis of classical music using the information length. *Entropy*, 18(7), pp.1–13.

Scolletta, S., Franchi, F., Romagnoli, S., Carlà, R., Donati, A., Fabbri, L.P., Forfori, F., Alonso-Iñigo, J.M., Laviola, S., Mangani, V., Maj, G., Martinelli, G., Mirabella, L., Morelli, A., Persona, P. and Payen, D., 2016. Comparison Between Doppler-Echocardiography and Uncalibrated Pulse Contour Method for Cardiac Output Measurement. *Critical Care Medicine*, [online] 44(7), pp.1370–1379.

Shrout, P.E. and Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), pp.420–428.

Schubert, D., Mühlstedt, J., Bullinger, A., 2015. *Usability Assessment of Medical Devices from a User Perspective: Results of a Focus Group Discussion*. In: Gurtner, S., Soyez, K. (eds) Challenges and Opportunities in Health Care Management. Springer, Cham.

Seidle, N., 2014. Arduino Scale With 5kg Load Cell and HX711 Amplifier. [Online] Available at: https://www.instructables.com/Arduino-Scale-With-5kg-Load-Cell-and-HX711-Amplifi/ [Accessed August 2017].

Shinners, P., 2011. PyGame - Python Game Development. Retrieved from: http://www.pygame.org

Shrout, P.E. and Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), pp.420–428.

Sigrist, R., Rauter, G., Riener, R. and Wolf, P., 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, [online] 20(1), pp.21–53.

Sihvonen, A.J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmüller, E. and Soinila, S., 2017. Music-based interventions in neurological rehabilitation. *The Lancet Neurology*, [online] 16(8), pp.648–660.

Sim, J. and Snell, J., 1996. Focus groups in physiotherapy evaluation and research. *Physiotherapy*, 82(3), pp.189–198.

Singh, A., Klapper, A., Jia, J., Fidalgo, A., Tajadura-Jiménez, A., Kanakam, N., Bianchi-Berthouze, N. and Williams, A., 2014. Motivating people with chronic pain to do physical activity. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [online] New York, NY, USA: ACM.pp.2803–2812. Available at: <http://dl.acm.org/citation.cfm?doid=2556288.2557268>.

Singh, A., Piana, S., Pollarolo, D., Volpe, G., Varni, G., Tajadura-Jiménez, A., Williams, A.C., Camurri, A. and Bianchi-Berthouze, N., 2016. Go-with-the-Flow : Tracking, Analysis and Sonification of Movement and Breathing to Build Confidence in Activity Despite Chronic Pain. *Human–Computer Interaction*, [online] 31(3–4), pp.335–383.

Singh, A., Bianchi-Berthouze, N. and Williams, A.C., 2017. Supporting Everyday Function in Chronic Pain Using Wearable Technology. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. [online] New York, NY, USA: ACM.pp.3903–3915.

Stark, T., Walker, B., Phillips, J.K., Fejer, R. and Beck, R., 2011. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: A systematic review. *PM and R*, [online] 3(5), pp.472–479.

Seitz, L.B. and Haff, G.G., 2016. Factors Modulating Post-Activation Potentiation of Jump, Sprint, Throw, and Upper-Body Ballistic Performances: A Systematic Review with Meta-Analysis. *Sports Medicine*, 46(2), pp.231–240.

Stevens, J.G.A. and Beurskens, A.J.M.H., 2010. Implementation of measurement instruments in physical therapist practice: Development of a tailored strategy. *Physical Therapy*, 90(6), pp.953–961.

Stockton, K., Wrigley, T., Mengersen, K., Kandiah, D., Paratz, J. and Bennell, K., 2011. Test–retest reliability of hand-held dynamometry and functional tests in systemic lupus erythematosus. *Lupus*, [online] 20(2), pp.144–150.

Stratford, P.W. and Balsor, B.E., 1994. A Comparison of Make and Break Tests Using a Hand-Held Dynamometer and the Kin-Com. *Journal of Orthopaedic & Sports Physical Therapy*, [online] 19(1), pp.28–32.

Sueyoshi, T., Nakahata, A., Emoto, G. and Yuasa, T., 2017. Single-Leg Hop Test Performance and Isokinetic Knee Strength After Anterior Cruciate Ligament Reconstruction in Athletes. *Orthopaedic Journal of Sports Medicine*, 5(11), pp.1–6.

Sung, K.-S., Yi, Y.G. and Shin, H.-I., 2019. Reliability and validity of knee extensor strength measurements using a portable dynamometer anchoring system in a supine position. *BMC Musculoskeletal Disorders*, [online] 20(1), p.320.

Swinkels, R.A.H.M., Van Peppen, R.P.S., Wittink, H., Custers, J.W.H. and Beurskens, A.J.H.M., 2011. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. *BMC Musculoskeletal Disorders*, [online] 12(1), p.106.

Tanaka, R., Takimoto, H., Yamasaki, T. and Higashi, A., 2018. Validity of time series kinematical data as measured by a markerless motion capture system on a flatland for gait assessment. *Journal of Biomechanics*, [online] 71, pp.281–285.

Tausch, A.P. and Menold, N., 2016. Methodological Aspects of Focus Groups in Health Research. *Global Qualitative Nursing Research*, [online] 3, pp.1–12.

Temple, M.D., 2017. An auditory display tool for DNA sequence analysis. *BMC Bioinformatics*, [online] 18(1), p.221.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M. and de Vet, H.C.W., 2007. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), pp.34–42.

Thomsen, A., 2014. *Tutorial: Acelerômetro MPU6050 com Arduino*. [Online] Available at: < https://www.filipeflop.com/blog/tutorial-acelerometro-mpu6050-arduino/

Thorborg, K., Petersen, J., Magnusson, S.P. and Hölmich, P., 2010. Clinical assessment of hip strength using a hand-held dynamometer is reliable. *Scandinavian Journal of Medicine and Science in Sports*, 20(3), pp.493–501.

Thorborg, K., Bandholm, T., & Hölmich, P., 2013. Hip-and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. *Knee Surgery, Sports Traumatology, Arthroscopy*, *21*(3), pp. 550-555.

Timmins, R.G., Shield, A.J., Williams, M.D. and Opar, D.A., 2016. Is There Evidence to Support the Use of the Angle of Peak Torque as a Marker of Hamstring Injury and Re-Injury Risk? *Sports Medicine*, 46(1), pp.7–13.

Tong, A., Sainsbury, P. and Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6), pp.349–357.

Toonstra, J. and Mattacola, C.G., 2013. Test-Retest Reliability and Validity of Isometric Knee-Flexion and -Extension Measurement Using 3 Methods of Assessing Muscle Strength. *Journal of Sport Rehabilitation*, [online] 22(1).

Trotter, M. and Gleser, G.C., 1958. A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology*, [online] 16(1), pp.79–123.

Turner, A.N. and Jeffreys, I., 2010. The stretch-shortening cycle: Proposed mechanisms and methods for enhancement. *Strength and Conditioning Journal*, 32(4), pp.87–99.

van der Meer, H.A., de Pijper, L., van Bruxvoort, T., Visscher, C.M., Nijhuis-van der Sanden, M.W.G., Engelbert, R.H. and Speksnijder, C.M., 2020. Using e-Health in the physical therapeutic care process for patients with temporomandibular disorders: a qualitative study on the perspective of physical therapists and patients. *Disability and Rehabilitation*, [online] 0(0), pp.1–8. Available at: <https://doi.org/10.1080/09638288.2020.1775900

Vanpee, G., Hermans, G., Segers, J. and Gosselink, R., 2014. Assessment of Limb Muscle Strength in Critically Ill Patients. *Critical Care Medicine*, [online] 42(3), pp.701–711.

Varni, G., Stoffregen, T., Camurri, A., Mazzarino, B. and Volpe, G., 2007. Audio Patterns as a Source of Information for the Perception and Control of Orientation. In: *2007 Virtual Rehabilitation*. [online] IEEE.pp.109–113.

Vermeulen, H.M., de Bock, G.H., van Houwelingen, H.C., van der Meer, R.L., Mol, M.C., Plus, B.T., Rozing, P.M. and Vliet Vlieland, T.P.M., 2005. A comparison of two portable dynamometers in the assessment of shoulder and elbow strength. *Physiotherapy*, 91(2), pp.101–112.

Vinken, P.M., Kröger, D., Fehse, U., Schmitz, G., Brock, H. and Effenberg, A.O., 2013. Auditory coding of human movement kinematics. *Multisensory Research*, [online] 26(6), pp.533–552.

Vogt, K., Pirrò, D., Kobenz, I., Höldrich, R. and Eckel, G., 2009. PhysioSonic – Movement sonification as auditory feedback. In: *Proceedings of the 15th International Conference on Auditory Display (ICAD 2009)*. [online] Copenhagen.pp.161–167.

Wadsworth, C.T., Krishnan, R., Sear, M., Harrold, J. and Nielsen, D.H., 1987. Intrarater reliability of manual muscle testing and hand-held dynametric muscle testing. *Physical Therapy*, 67(9), pp.1342–1347.

Walker, B.N. and Nees, M.A., 2011. Theory of sonification. In: T. Hermann, A. Hunt and J. Neuhoff, ed. 2011. *The sonification handbook*. Logos Verlag, Berlin, Germany. Ch. 2.

Wallis, R.I., Ingalls, T., Rikakis, T., Olsen, L., Chen, Y., Xu, W. and Sundaram, H., 2007. Real-time sonification movement for an immersive stroke rehabilitation environment. *Proceedings of the 13th International Conference on Auditory Display (ICAD 2007)*, [online] pp.497–503. Available at: <http://www.icad.org/node/2460>.

Wang, Q., Turaga, P., Coleman, G. and Ingalls, T., 2014. SomaTech: An Exploratory Interface for Altering Movement Habits. In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. [online] New York, NY, USA: ACM.pp.1765–1770.

WCPT. 2019. Policy statement: Description of physical therapy. http://www.wcpt.org/policy/ps-descriptionPT#appendix_1

Weir, J.P., 2005. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *The Journal of Strength and Conditioning Research*, [online] 19(1), pp.231–240.

Wellsandt, E., Failla, M.J. and Snyder-Mackler, L., 2017. Limb symmetry indexes can overestimate knee function after anterior cruciate ligament injury. *Journal of Orthopaedic and Sports Physical Therapy*, 47(5), pp.334–338.

Whelan, D.F., O'Reilly, M.A., Ward, T.E., Delahunt, E. and Caulfield, B., 2017. Technology in Rehabilitation: Evaluating the Single Leg Squat Exercise with Wearable Inertial Measurement Units. *Methods of Information in Medicine*, [online] 56(02), pp.88–94.

Whinton, A.K., Thompson, K.M.A., Power, G.A. and Burr, J.F., 2018. Testing a novel isokinetic dynamometer constructed using a 1080 Quantum. *PLoS ONE*, 13(7), pp.1–11.

Whiteley, R., Jacobsen, P., Prior, S., Skazalski, C., Otten, R. and Johnson, A., 2012. Correlation of isokinetic and novel hand-held dynamometry measures of knee flexion and extension strength testing. *Journal of Science and Medicine in Sport*, [online] 15(5), pp.444–450.

Whiteley, R., van Dyk, N., Wangensteen, A. and Hansen, C., 2018. Clinical implications from daily physiotherapy examination of 131 acute hamstring injuries and their association with running speed and rehabilitation progression. *British Journal of Sports Medicine*, 52(5), pp.303–310.

Wikholm, J.B. and Bohannon, R.W., 1991. Hand-held dynamometer measurements: Tester strength makes a difference. *Journal of Orthopaedic and Sports Physical Therapy*, 13(4), pp.191–198.

Winkelman, N.C., Clark, K.P. and Ryan, L.J., 2017. Experience level influences the effect of attentional focus on sprint performance. *Human Movement Science*, [online] 52, pp.84–95.

Winton, R.J., Gable, T.M., Schuett, J. and Walker, B.N., 2012. A sonification of Kepler space telescope star data. *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012)*, pp.218–220.

Wilhite, M.R., Cohen, E.R. and Wilhite, S.C., 1992. Reliability of concentric and eccentric measurements of quadriceps performance using the KIN-COM

dynamometer: The effect of testing order for three different speeds. *Journal of Orthopaedic and Sports Physical Therapy*, 15(4), pp.175–182.

Wintz, M.M., 1959. Variations in current manual muscle testing. *The Physical therapy review*, 39(7), pp.466–475.

Wyrwich, K., Nienaber, N., Tierney, W., and Wolinsky, F., 1999. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Medical care*, 37(5), pp.469–478

Wolf, P., Sigrist, R., Rauter, G. and Riener, R., 2011. Error Sonification of a Complex Motor Task. *BIO Web of Conferences*, [online] 1, pp.1–4.

Wolfe, J., 1997. Note names, MIDI numbers and frequencies. [Website] Available at: https://newt.phys.unsw.edu.au/jw/notes.html [Accessed 16 April 2020]

Wollin, M., Thorborg, K., Drew, M. and Pizzari, T., 2020. A novel hamstring strain injury prevention system: Post-match strength testing for secondary prevention in football. *British Journal of Sports Medicine*, 54(9), pp.1–2.

Wright, W.G., 1912. Muscle Training in the Treatment of Infantile Paralysis. *The Boston Medical and Surgical Journal*, [online] 167(17), pp.567–574.

Wulf, G., 2013. Attentional focus and motor learning: A review of 15 years. *International Review of Sport and Exercise Psychology*, 6(1), pp.77–104.

Zaina, L.A.M. and Álvaro, A., 2015. A design methodology for user-centered innovation in the software development area. *Journal of Systems and Software*, [online] 110, pp.155–177.

Zaki, R., Bulgiba, A., Ismail, R. and Ismail, N.A., 2012. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. *PLoS ONE*, 7(5), pp.1–7.

Zatsiorsky, V. M., and Kraemer, W. J., 2020. *Science and practice of strength training*. Champaign, IL: Human Kinetics.

APPENDICES

Appendix 1

Appendix 1. Description of Randomised Controlled Trials

Author(s)	Country	Study Outline	Sample	Intervention	Key Findings
Effenberg, 2005	Germany	Assess perception and action, subjects to estimate and reproduce countermovement jump using video and sonified feedback.	Two exp. – 2 groups 40 healthy subjects	Exp. 1 - Movement assessment of 2 consecutive countermovement jumps. Exp. 2 - Subjects observed reproduction of a single jump of different heights.	Exp. 1 - Movement sonification helps improve perception accuracy of sonified sports movements; Exp. 2 - Additional convergent auditory stimulus can enhance sports movements' reproduction accuracy compared to the video feedback.

Giansanti et al., 2009	Italy	Assess trunk postural movement induced by audio feedback using a wearable device.	Nine healthy subjects	Five trials of 50s each in 6 different conditions: 1-eyes closed on a solid surface; 2- eyes open on foam cushion; 3- eyes closed on foam cushion; 4- Eyes closed with audio feedback; 5- eyes open on foam with audio-feedback; 6- eyes closed on foam with audio-feedback.	Subjects able to save energy while using the audio-feedback system. Sonification facilitates postural control in conditions with reduced sensorial input.
Pauletto and Hunt, 2009	United Kingdom	Two exp. using large datasets: explore interactive sonification with recorded data and real-time data (from EMG sensors)	Two Exp.: 1-21 2-57 (17 with OA)	1- Subjects listened to sonified datasets; 2- EMG data sonified, users, asked for feedback about roughness, overall loudness, speed of sounds' attack.	Complex data can be used in sonification especially if the user can interact with the sound display. The system might also be used as a rehabilitation tool in the future.
Dailly, et al., 2012	Switzerla nd	Assess learning accuracy of a new task using error	12 healthy subjects	Subjects traced a pre-defined trajectory on a table surface.	Concurrent error sonification appears to be more beneficial than repetitive

		sonification and music in an upper limb movement			training in healthy subjects learning a new upper limb movement.
Vinken, et al., 2013	Germany	Investigation on types of sonification use for upper limb gross movement	28 healthy subjects	Individuals listened to sonification of human movements 126 times in total. 3 times to 6 upper limb actions sonified in 7 different ways.	Sonification might be used for motor control and rehabilitation as it improves the perception of human movement even in subjects not familiar with sonification.
Scholz, et al., 2014	Germany	Develop a sonification based stroke rehab protocol that provides additional sensory input in upper limb movements.	26 healthy subjects	Subjects randomised to start the procedure with one of two different conditions which had different grid orientation on brightness and pitch. Two conditions were tested. The pitch and brightness representation were swapped for the second condition.	Pitch benefited from being set on the vertical axis and brightness on the horizontal axis. This was done using only bi-dimensional sonification.

Schmitz, Kroeger and Effenberg, 2014	Germany	Develop a sonification based mobile rehabilitation system to support stroke patients.	Seven patients	Subjects performed upper limb movements of different complexity for five days with each session up to 20 minutes each. 4 of the subjects were included in the sonification group while 3 were part of the control group.	Improvement on gross motor skills, although the sample was small and heterogeneous. However, the system developed provides a good example of how these can be built.
Scholz, et al., 2015	Germany	To develop a solution for stroke rehabilitation using music sonification therapy.	Four stroke patients	Two groups (2 patients each) received nine days of music sonification therapy or sham sonification training.	Music sonification therapy appears to have beneficial results as it is highly motivating and can support sensorial input affected by stroke. Subjects improved more than the control group in the motor function test.
Fujii, Lulic and Chen, 2016	Canada	Subjects asked to hit a target in an upper limb movement sonification task.	20 healthy Subjects	All the participants had 25 trials to learn a "reaching movement". One group received sonification in 100% of the trials (100 times), and a second group exposed to sonification in 50% of the trials (100 times). Both groups had 25 testing trials to assess retention on day	Upper limb complex motor tasks might benefit from more concurrent knowledge of performance audio-feedback (100% of the time) than (50% of the time) of the trials while learning a movement pattern.

				one, and on day two, with 25 more trials, no feedback was given in the retention tests.	
Newbold, et al., 2016	United Kingdom	Technical solution of mobile sonification to facilitate and guide movement in patients with chronic pain.	17 healthy Subjects	Smartphone attached to subjects' trunk to analyse trunk displacement (forward) and provide audio feedback about movement quantity, time of return, and self-reported measures.	This study shows that music-based sonifications can be used to provide feedback on the range of movement and how it can be perceived as more rewarding by healthy subjects.
Scholz, et al., 2016	Germany	Music sonification therapy for stroke patient's rehabilitation	25 stroke patients	Patients with moderate impairment in upper limb motor function randomly assigned in two groups that received an average of 10 days sonification therapy or sonification sham. Assessed pre and post-training in different parameters (upper extremity	This type of sonification in the rehabilitation of stroke patients seems to be beneficial due to its improvements in gross motor function and decreased joint pain.

				function; psychological state and arm movement smoothness).	
Dyer, Stapleton and Rodger, 2017	UK	To investigate the advantages of melodic sonification in a new bi-manual motor skill task.	60 healthy subjects	Three groups with three different sound feedback, randomly distributed. Terminal feedback was provided to all groups (graph). 1 - Control 2 - Melodic sonification 3 - Rhythmic sonification.	Melodic sonification was more effective than rhythmic and control group for motor task learning. There was no significant difference between the control group and the rhythmic sonification which was unexpected.
Newbold, Bianchi-Berthouze and Gould, 2017	UK	To exploit music's qualities in a squat sonification task	20 healthy subjects in each experiment	Participants were asked to squat while using a mobile smartphone attached to the thigh. First exp. they explored musical expectancy; Second exp. subjects squatted (once) at a steady pace to each chord until the end of the movement while hearing the "stable" or "unstable" sonification only.	1- Results show differences between the musical sonifications with the non-musical/no sound, with more movement past the target point, slower return and better motivation. 2- No significant effects on additional squat or time of return. However, in the "stable" condition, individuals thought they had moved more than

					they really did, while in the "unstable" condition, individuals felt more motivated to continue the movement.
Alcaraz, et al. (2018)	Germany	To use machine learning to develop sonification for a gait training task with healthy subjects	20 Healthy participants	Machine learning sonification was played to the participants in the experimental group (n=10)	Improved results in the sonification group in terms of gait speed, cadence and stride length.
Bevilacqua et al., 2018	France	To investigate how movement sonification can be used for stroke patients when using	8 stroke patients and 7 healthy individuals	Different sound-feedback was played to patients and healthy individuals, such as direct sonification, musical interaction, environmental sounds exploration	Direct mapping was the least preferred by the stroke patients, with authors pointing that it appears that individuals either healthy or post-

		sonification for upper limb movements		stroke have different selections of preferred sound-feedback.
Maes, Lorenzoni & Six, 2018	Belgium	To assess the ability of a musical sonification to increase synchronization on a static bike movement	15 healthy participants	Five different sound conditions were used on participants using a stationary bike. Three different music sonification models were used plus two no sonification conditions (no sound and non-sonified feedback)	Participants tended to synchronize their rhythm and cadence was more stable on sonified musical conditions than the non-sonified condition. Authors also state this is a preliminary work and further work is needed to confirm these conclusions.

| Nikmaram, et al., 2019 | Germany | 1 – Replicate results from previous pilot 2 – Test a less costly solution for hand-tracking 3 – Estimate efficiency of musical sonification in stroke patients by comparing the results from both data collections | Acute or subacute unilateral stroke patients 1 – 12 2 - 30 | Sonification of upper limb movements in addition to regular physiotherapy intervention for the intervention group | Limited benefits arose from the sonification (on movement smoothness), authors recommend it being used as an adjunct to common approaches. Authors conclude that patient "enthusiasm" was relevant but do not report it as outcome. |
| **Reh, et al., 2019** | Germany | Investigate the use of sonification in gait symmetry ad steadiness in patients with unilateral hip arthroplasty | 20 patients | 10 patients for either control group (20 minute gait training with no sound-feedback) or intervention group (20 minute gait training with sound-feedback) | Step length and symmetry converged on the sonification group but not on the control group, however the authors state that larger variability on those parameters were more evident on the |

				intervention group than the control group
Haire, et al. 2021	Canada	Assess the effect of sonification and rhythmic auditory stimulation approaches in chronic post-stroke patients in upper limb movements	30 chronic post-stroke participants	A three-armed parallel RCT which compared the effects of audio-feedback under three conditions (alone, with motor imagery and motor imagery with audio-cueing) on the Fugl-Meyer–Upper Extremity and Wolf Motor Function Test–Functional Ability Scale	Improvement in the assessed outcomes which suggests this could be a useful approach for chronic stroke patients, sample sizes were small and heterogenous which could limit results extrapolation.
O'Brien, et al., 2021	France	To assess the use of sonification in putting and swing	40 healthy individuals	A four group intervention program was designed with movement analysed in two different conditions (2m and 4m putts). Data	Error-based sonification group benefited more in reducing variability and timing of the putting motions

		movement in golf novices		were compared for the control group vs the three audio-feedback groups	when compared to the remaining intervention groups.
Mezzaroba , et al., 2020	Italy	Investigate the use of action observation protocol in	22 Parkinson's Disease patients	Participants were exposed to videos with eight motor gestures either in the sonification group or in the control group. All interventions were assessed before and after intervention using a sit-to-walk task	Results show an improvement on the freezing of gait of patients in the intervention group using sonification which authors reporting better dynamic balance control.

Note: Exp. – Experiment; OA – Osteoarthritis; s - seconds

Appendix 2

Arduino UNO R3 datasheet

Retrieved from: https://store.arduino.cc/arduino-uno-rev3

OVERVIEW | TECH SPECS | DOCUMENTATION | FAQ

Microcontroller	ATmega328P
Operating Voltage	5V
Input Voltage (recommended)	7-12V
Input Voltage (limit)	6-20V
Digital I/O Pins	14 (of which 6 provide PWM output)
PWM Digital I/O Pins	6
Analog Input Pins	6
DC Current per I/O Pin	20 mA
DC Current for 3.3V Pin	50 mA
Flash Memory	32 KB (ATmega328P) of which 0.5 KB used by bootloader
SRAM	2 KB (ATmega328P)
EEPROM	1 KB (ATmega328P)
Clock Speed	16 MHz
LED_BUILTIN	13
Length	68.6 mm
Width	53.4 mm
Weight	25 g

Appendix 3

MPU6050 - Datasheet

This annex is part of the full document available at:

https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf

6   Electrical Characteristics

6.1   Gyroscope Specifications

VDD = 2.375V-3.46V, VLOGIC (MPU-6050 only) = 1.8V±5% or VDD, T$_A$ = 25°C

PARAMETER	CONDITIONS	MIN	TYP	MAX	UNITS	NOTES
GYROSCOPE SENSITIVITY						
Full-Scale Range	FS_SEL=0		±250		°/s	
	FS_SEL=1		±500		°/s	
	FS_SEL=2		±1000		°/s	
	FS_SEL=3		±2000		°/s	
Gyroscope ADC Word Length			16		bits	
Sensitivity Scale Factor	FS_SEL=0		131		LSB/(°/s)	
	FS_SEL=1		65.5		LSB/(°/s)	
	FS_SEL=2		32.8		LSB/(°/s)	
	FS_SEL=3		16.4		LSB/(°/s)	
Sensitivity Scale Factor Tolerance	25°C	-3		+3	%	
Sensitivity Scale Factor Variation Over Temperature			±2		%	
Nonlinearity	Best fit straight line; 25°C		0.2		%	
Cross-Axis Sensitivity			±2		%	
GYROSCOPE ZERO-RATE OUTPUT (ZRO)						
Initial ZRO Tolerance	25°C		±20		°/s	
ZRO Variation Over Temperature	-40°C to +85°C		±20		°/s	
Power-Supply Sensitivity (1-10Hz)	Sine wave, 100mVpp; VDD=2.5V		0.2		°/s	
Power-Supply Sensitivity (10 - 250Hz)	Sine wave, 100mVpp; VDD=2.5V		0.2		°/s	
Power-Supply Sensitivity (250Hz - 100kHz)	Sine wave, 100mVpp; VDD=2.5V		4		°/s	
Linear Acceleration Sensitivity	Static		0.1		°/s/g	
SELF-TEST RESPONSE						
Relative	Change from factory trim	-14		14	%	1
GYROSCOPE NOISE PERFORMANCE	FS_SEL=0					
Total RMS Noise	DLPFCFG=2 (100Hz)		0.05		°/s-rms	
Low-frequency RMS noise	Bandwidth 1Hz to10Hz		0.033		°/s-rms	
Rate Noise Spectral Density	At 10Hz		0.005		°/s/ √ Hz	
GYROSCOPE MECHANICAL FREQUENCIES						
X-Axis		30	33	36	kHz	
Y-Axis		27	30	33	kHz	
Z-Axis		24	27	30	kHz	
LOW PASS FILTER RESPONSE						
	Programmable Range	5		256	Hz	
OUTPUT DATA RATE						
	Programmable	4		8,000	Hz	
GYROSCOPE START-UP TIME	DLPFCFG=0					
ZRO Settling (from power-on)	to ±1°/s of Final		30		ms	

1.   Please refer to the following document for further information on Self-Test: *MPU-6000/MPU-6050 Register Map and Descriptions*

6.2 Accelerometer Specifications

VDD = 2.375V-3.46V, VLOGIC (MPU-6050 only) = 1.8V±5% or VDD, T_A = 25°C

PARAMETER	CONDITIONS	MIN	TYP	MAX	UNITS	NOTES
ACCELEROMETER SENSITIVITY						
Full-Scale Range	AFS_SEL=0		±2		g	
	AFS_SEL=1		±4		g	
	AFS_SEL=2		±8		g	
	AFS_SEL=3		±16		g	
ADC Word Length	Output in two's complement format		16		bits	
Sensitivity Scale Factor	AFS_SEL=0		16,384		LSB/g	
	AFS_SEL=1		8,192		LSB/g	
	AFS_SEL=2		4,096		LSB/g	
	AFS_SEL=3		2,048		LSB/g	
Initial Calibration Tolerance			±3		%	
Sensitivity Change vs. Temperature	AFS_SEL=0, -40°C to +85°C		±0.02		%/°C	
Nonlinearity	Best Fit Straight Line		0.5		%	
Cross-Axis Sensitivity			±2		%	
ZERO-G OUTPUT						
Initial Calibration Tolerance	X and Y axes		±50		mg	1
	Z axis		±80		mg	
Zero-G Level Change vs. Temperature	X and Y axes, 0°C to +70°C		±35			
	Z axis, 0°C to +70°C		±60		mg	
SELF TEST RESPONSE						
Relative	Change from factory trim	-14		14	%	2
NOISE PERFORMANCE						
Power Spectral Density	@10Hz, AFS_SEL=0 & ODR=1kHz		400		µg/ √Hz	
LOW PASS FILTER RESPONSE						
	Programmable Range	5		260	Hz	
OUTPUT DATA RATE						
	Programmable Range	4		1,000	Hz	
INTELLIGENCE FUNCTION INCREMENT			32		mg/LSB	

1. Typical zero-g initial calibration tolerance value after MSL3 preconditioning
2. Please refer to the following document for further information on Self-Test: *MPU-6000/MPU-6050 Register Map and Descriptions*

292

6.3 Electrical and Other Common Specifications

VDD = 2.375V-3.46V, VLOGIC (MPU-6050 only) = 1.8V±5% or VDD, T_A = 25°C

PARAMETER	CONDITIONS	MIN	TYP	MAX	Units	Notes
TEMPERATURE SENSOR						
Range			-40 to +85		°C	
Sensitivity	Untrimmed		340		LSB/°C	
Temperature Offset	35°C		-521		LSB	
Linearity	Best fit straight line (-40°C to +85°C)		±1		°C	
VDD POWER SUPPLY						
Operating Voltages		2.375		3.46	V	
Normal Operating Current	Gyroscope + Accelerometer + DMP		3.9		mA	
	Gyroscope + Accelerometer (DMP disabled)		3.8		mA	
	Gyroscope + DMP (Accelerometer disabled)		3.7		mA	
	Gyroscope only (DMP & Accelerometer disabled)		3.6		mA	
	Accelerometer only (DMP & Gyroscope disabled)		500		µA	
Accelerometer Low Power Mode Current	1.25 Hz update rate		10		µA	
	5 Hz update rate		20		µA	
	20 Hz update rate		70		µA	
	40 Hz update rate		140		µA	
Full-Chip Idle Mode Supply Current			5		µA	
Power Supply Ramp Rate	Monotonic ramp. Ramp rate is 10% to 90% of the final value			100	ms	
VLOGIC REFERENCE VOLTAGE	MPU-6050 only					
Voltage Range	VLOGIC must be ≤VDD at all times	1.71		VDD	V	
Power Supply Ramp Rate	Monotonic ramp. Ramp rate is 10% to 90% of the final value			3	ms	
Normal Operating Current			100		µA	
TEMPERATURE RANGE						
Specified Temperature Range	Performance parameters are not applicable beyond Specified Temperature Range	-40		+85	°C	

Appendix 4

HX-711 Analog to Digital Converter

This factsheet is part of the full documentation available at:
https://cdn.sparkfun.com/datasheets/Sensors/ForceFlex/hx711_english.pdf

海芯科技 AVIA Semiconductor

HX711

24-Bit Analog-to-Digital Converter (ADC) for Weigh Scales

DESCRIPTION

Based on Avia Semiconductor's patented technology, HX711 is a precision 24-bit analog-to-digital converter (ADC) designed for weigh scales and industrial control applications to interface directly with a bridge sensor.

The input multiplexer selects either Channel A or B differential input to the low-noise programmable gain amplifier (PGA). Channel A can be programmed with a gain of 128 or 64, corresponding to a full-scale differential input voltage of $\pm20mV$ or $\pm40mV$ respectively, when a 5V supply is connected to AVDD analog power supply pin. Channel B has a fixed gain of 32. On-chip power supply regulator eliminates the need for an external supply regulator to provide analog power for the ADC and the sensor. Clock input is flexible. It can be from an external clock source, a crystal, or the on-chip oscillator that does not require any external component. On-chip power-on-reset circuitry simplifies digital interface initialization.

There is no programming needed for the internal registers. All controls to the HX711 are through the pins.

FEATURES

- **Two selectable differential input channels**
- **On-chip active low noise PGA with selectable gain of 32, 64 and 128**
- **On-chip power supply regulator for load-cell and ADC analog power supply**
- **On-chip oscillator requiring no external component with optional external crystal**
- **On-chip power-on-reset**
- **Simple digital control and serial interface:**
  pin-driven controls, no programming needed
- **Selectable 10SPS or 80SPS output data rate**
- **Simultaneous 50 and 60Hz supply rejection**
- **Current consumption including on-chip analog power supply regulator:**
  normal operation < 1.5mA, power down < 1uA
- **Operation supply voltage range: 2.6 ~ 5.5V**
- **Operation temperature range: –40 ~ +85℃**
- **16 pin SOP-16 package**

APPLICATIONS

- **Weigh Scales**
- **Industrial Process Control**
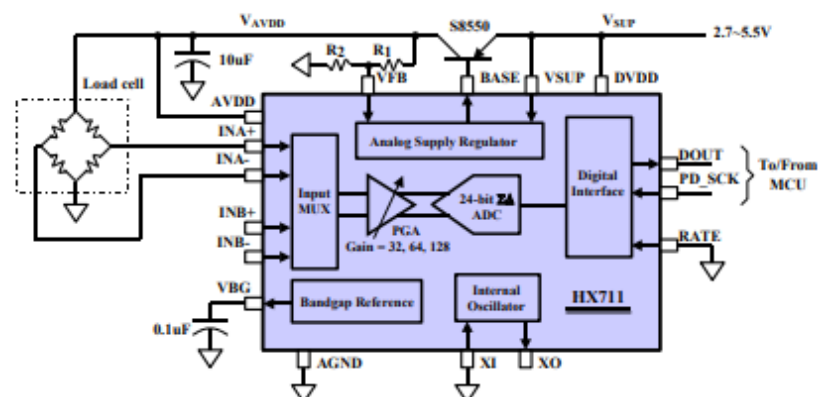
Fig. 1 Typical weigh scale application block diagram

KEY ELECTRICAL CHARACTERISTICS

Parameter	Notes	MIN	TYP	MAX	UNIT
Full scale differential input range	V(inp)-V(inn)		± 0.5(AVDD/GAIN)		V
Common mode input		AGND+1.2		AVDD-1.3	V
Output data rate	Internal Oscillator, RATE = 0		10		Hz
	Internal Oscillator, RATE = DVDD		80		
	Crystal or external clock, RATE = 0		f_{clk}/1,105,920		
	Crystal or external clock, RATE = DVDD		f_{clk}/138,240		
Output data coding	2's complement	800000		7FFFFF	HEX
Output settling time [1]	RATE = 0		400		ms
	RATE = DVDD		50		
Input offset drift	Gain = 128		0.2		mV
	Gain = 64		0.4		
Input noise	Gain = 128，RATE = 0		50		nV(rms)
	Gain = 128，RATE = DVDD		90		
Temperature drift	Input offset（Gain = 128）		± 6		nV/℃
	Gain（Gain = 128）		± 5		ppm/℃
Input common mode rejection	Gain = 128，RATE = 0		100		dB
Power supply rejection	Gain = 128，RATE = 0		100		dB
Reference bypass（V_{BG}）			1.25		V
Crystal or external clock frequency		1	11.0592	20	MHz
Power supply voltage	DVDD	2.6		5.5	V
	AVDD，VSUP	2.6		5.5	
Analog supply current (including regulator)	Normal		1400		μA
	Power down		0.3		
Digital supply current	Normal		100		μA
	Power down		0.2		

（1）Settling time refers to the time from power up, reset, input channel change and gain change to valid stable output data.

Table 2 Key Electrical Characteristics

Appendix 4

PSD-S1 Load Cell

Load cell specifications:

Model: PSD-S1

Output signal: analog

Linearity: ±0.02（%F.S.）

Delaying: ±0.02（%F.S.）

Repeatability: ±0.02（%F.S.）

Sensitivity: 2.0mv/v

Drifting: ±0.02%F.S

Input impedance: 350±10Ω

Output impedance: 350±5Ω

Measuring range:300kg

Weight: 575g

Wire connecting method:

Input: Red wire(+), Black wire(-)

Output: Green wire(+), White wire(-)

Yellow wire(GND)

Package Included:

1 x Weighting Sensor

Appendix 6

Hypothesis

General Hypothesis

1- Intra-tester reliability

a) Peak force

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis - A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) Angle of Peak force

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

c) Angular impulse

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

2 - Inter-tester reliability

a) Peak force

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) Angle of Peak force

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

c) **Angular impulse**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

3 - Validity

a) **Peak torque**

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

b) **Angle of Peak torque**

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

4- Sonification testing

a) **Intra-tester reliability**

Null hypothesis – – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests does not increase reliability testing in a healthy population.

Alternative hypothesis – The use of sonification physiotherapy by students and experienced physiotherapists in concentric manual muscle tests increases reliability testing in a healthy population.

b) **Inter-tester reliability**

Null hypothesis – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests does not increase reliability testing in a healthy population.

Alternative hypothesis – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests increases reliability testing in a healthy population.

c) **Force production**

Null hypothesis – The use of sonification in concentric manual muscle tests does not increase force production in physiotherapy students and experienced physiotherapists in a healthy population.

Alternative hypothesis – The use of sonification in concentric manual muscle tests increases force production in physiotherapy students and experienced physiotherapists in a healthy population.

Hypothesis for chapter 4

Below are the hypothesis tested for both movements of knee extension and knee flexion and for each of the testers as appropriate.

1. **Intra-tester reliability**
a) **Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis - A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) **Angle of Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

2 - Inter-tester reliability

a) **Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) **Angle of Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

3 - Validity

a) **Peak torque**

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

b) **Angle of Peak torque**

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

The null hypothesis 1a for peak force can be rejected and alternative hypothesis 1a accepted with ASSA v2 with a reliability above 0.75 for both movements. The null hypothesis 1b for angle of peak force reliability is accepted, with the device below the minimum acceptable levels. In terms of inter-tester reliability, the null hypothesis can also be accepted for both hypothesis 2a and 2b. Data from the validity analysis allows the author to reject the null hypothesis for 3a but not for 3b, with the device able to detect peak torque but not angle of peak torque.

Hypothesis for chapter 5

1 - Intra-tester reliability

a) **Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis - A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) **Angle of Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

c) **Angular impulse**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

2- Inter-tester reliability

a) Peak force

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects.

b) Angle of Peak force

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

c) Angular impulse

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects.

3- Validity

a) Peak torque

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

b) **Angle of Peak torque**

Null hypothesis – A concentric MMT using ASSA is not a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA is a valid device to detect angle of peak torque when compared to current gold standards (isokinetic dynamometer) in healthy subjects.

The null hypothesis 1a can be rejected for in all movements for tester one and in tester 2 for all movements except for knee flexion. In contrast, the null hypothesis 1b can not be rejected for either tester. In terms of the null hypothesis 1c, it can be rejected for tester 2 in the shoulder and elbow movements and for tester 1 in elbow flexion, but not for the remaining movements tested.

The null hypothesis 2a can be rejected for all movements except knee flexion. At the same time, the null hypothesis 2b can be rejected for elbow flexion. The null hypothesis 2c can be rejected only for shoulder abduction and elbow flexion.

The null hypothesis regarding the validity of peak force 3a can be rejected, but the null hypothesis 3b can not be rejected as the device is not valid to detect the angle of peak torque.

Hypothesis for chapter 6

1 **- Intra-tester reliability**

a) **Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects in testers of different sex and experience.

Alternative hypothesis - A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect peak force in healthy subjects in testers of different sex and experience.

b) **Angle of Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects in testers of different sex and experience.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects in testers of different sex and experience.

c) **Angular impulse**

Null hypothesis – A concentric MMT using ASSA does not have good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects in testers of different sex and experience.

Alternative hypothesis – A concentric MMT using ASSA has good intra-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects in testers of different sex and experience.

2 **- Inter-tester reliability**

a) **Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects in testers of different sex and experience.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect peak force in healthy subjects in testers of different sex and experience.

b) **Angle of Peak force**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angle of peak force in healthy subjects in testers of different sex and experience.

c) **Angular impulse**

Null hypothesis – A concentric MMT using ASSA does not have good inter-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects in testers of different sex and experience.

Alternative hypothesis – A concentric MMT using ASSA has good inter-tester reliability (ICC>0.75) to detect angular impulse in healthy subjects in testers of different sex and experience.

3- Sonification testing

a) **Intra-tester reliability**

Null hypothesis – – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests does not increase reliability in a healthy population.

Alternative hypothesis – The use of sonification physiotherapy by students and experienced physiotherapists in concentric manual muscle tests increases reliability in a healthy population.

b) **Inter-tester reliability**

Null hypothesis – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests does not increase reliability in a healthy population.

Alternative hypothesis – The use of sonification by physiotherapy students and experienced physiotherapists in concentric manual muscle tests increases reliability in a healthy population.

c) **Force production**

Null hypothesis – The use of sonification in concentric manual muscle tests does not increase force production in physiotherapy students and experienced physiotherapists in a healthy population.

Alternative hypothesis – The use of sonification in concentric manual muscle tests increases force production in physiotherapy students and experienced physiotherapists in a healthy population.

In terms of intra-tester reliability, even with testers with different experience, the null hypothesis (1a and 1c) can be rejected for peak force and angular impulse but not for angle of peak force (hypothesis 1b). Whereas for inter-tester reliability, the null hypothesis (2a) can be rejected for the use of concentric tests for elbow and knee flexion (ICC >0.75) (p<0.001) peak force detection, but not for the angle of peak force (2b). The angular impulse null hypothesis can be rejected (2c) for elbow flexion (ICC>0.75 p<0.001) but not for knee flexion (ICC<0.75).

For the hypothesis regarding the sonification testing, the null hypothesis 3a referring for the effects of sonification in intra-tester reliability can be rejected for elbow flexion peak force in all but one tester, whereas it can be fully rejected for knee flexion peak force. In terms of angular impulse the null hypothesis can also be rejected in angular impulse for elbow flexion in all testers except tester 4, whereas in angular impulse for knee flexion the null hypothesis 3a can be rejected only for two testers. The null hypothesis, 3b, can be rejected for peak force with sonification improving reliability for both movements tested and angular impulse for elbow flexion but not for knee flexion. Lastly, it cannot be rejected for angle of peak force in either joint. For the last hypothesis, 3c, the null hypothesis can be rejected only for tester 4 on both joints and for tester 3 on knee flexion.

Appendix 7

Ethics approval – ESPGR-02

	Joao Guerra
Project title:	Development of a portable strength assessment device for human movement via sonification – Arduino-based Sound Strength Assessment (ASSA)
DREP code:	ESPGR-02
Approval date	Valid to 25/9/2021

Application decision: **Approve** under the terms of Anglia Ruskin University's Research Ethics Policy (Dated 8 September 2016, Version 1.7). Approval by DREP is subject to ratification by the FREP.

Ethical approval is given for a period of 3 years doctorate students. If your research will extend beyond this period, it is your responsibility to apply for an extension before your approval expires.

It is your responsibility to ensure that you comply with Anglia Ruskin University's Research Ethics Policy and the Code of Practice for Applying for Ethical Approval at Anglia Ruskin University available at www.anglia.ac.uk/researchethics including the following.

- The procedure for submitting substantial amendments to the committee, should there be any changes to your research. You cannot implement these amendments until you have received approval from DREP for them.
- The procedure for reporting accidents, adverse events and incidents.
- The Data Protection Act (1998) and General Data Protection Requirement from 25 May 2018.
- Any other legislation relevant to your research. You must also ensure that you are aware of any emerging legislation relating to your research and make any changes to your study (which you will need to obtain ethical approval for) to comply with this.
- Obtaining any further ethical approval required from the organisation or country (if not carrying out research in the UK) where you will be carrying the research out. This includes other Higher Education Institutions if you intend to carry out any research involving their students, staff or premises. Please ensure that you send the DREP copies of this documentation if required, prior to starting your research.
- Any laws of the country where you are carrying the research and obtaining any other approvals or permissions that are required.

- Any professional codes of conduct relating to research or requirements from your funding body (please note that for externally funded research, where the funding has been obtained via Anglia Ruskin University, a Project Risk Assessment must have been carried out prior to starting the research).
- Completing a Risk Assessment (Health and Safety) if required and updating this annually or if any aspects of your study change which affect this.
- Notifying the DREP Secretary when your study has ended.

Please also note that your research may be subject to monitoring.

Should you have any queries, please do not hesitate to contact me. May I wish you the best of luck with your research.

Yours sincerely,

DREP Chair

Date 6.10.17

V1.2

Appendix 8

Ethics approval – ESPGR-05

Principal investigator:	Joao Guerra
Project supervisor:	Lee Smith
Project title:	The potential yield of external feedback via movement sonification in physiotherapy – Sonification mapping assessment
SREP code:	ESPGR-05
Approval date	Valid to 7/5/2021

Application decision: **Approve with revisions** under the terms of Anglia Ruskin University's Research Ethics Policy (Dated 8 September 2016, Version 1.7). Approval by SREP is subject to ratification by the FREP.

Changes to be made: These changes should be discussed and approved by your supervisor.

All recommendations must be completed and a copy of your updated submission sent to sportandexercisesciences-DREP@anglia.ac.uk by your supervisor as soon as possible.

GENERAL COMMENTS FOR THE APPLICANT	Please review specific comments below to ensure the application form and the PIS provide appropriate and consistent information. Risk assessment to conduct work in the lab needed.
Specific Comments for applicant: **Application Form**	
Section 1	Approved
Section 2	You answered 'No' to Q5, but then discuss gatekeepers in Section 4. Suggest reviewing and providing a gatekeeper email template.
Section 3	Approved
Section 4	Please, clarify which of your answers refer to each of the specific questions in Section 2. You answered 'Yes' to three points, but provide four answers in Section 4.
Section 5	Approved

Specific Comments for applicant: **PIS**	
Section A	Review the document to ensure correct grammar and appropriate language is used as this is a document shared with external audience. Point 5 should provide specific information on inclusion criteria as reported in Stage 1 form.
Section B	Review the document to ensure correct grammar and appropriate language is used as this is a document shared with external audience.
Specific Comments for: **Other Documentation**	
Consent Form	Approved
Risk Assessment	Missing
Ethics Quiz	Missing (assume this has been approved before, as this is a PhD student)
Pre-Exercise Questionnaire	Missing – is this needed?
Include missing documents:	Gatekeeper email is required

Ethical approval is given for a period of 1 year for undergraduates/masters students. If your research will extend beyond this period, it is your responsibility to apply for an extension before your approval expires.

It is your responsibility to ensure that you comply with Anglia Ruskin University's Research Ethics Policy and the Code of Practice for Applying for Ethical Approval at Anglia Ruskin University available at www.anglia.ac.uk/researchethics including the following.

- The procedure for submitting substantial amendments to the committee, should there be any changes to your research. You cannot implement these amendments until you have received approval from FREP/SREP for them.
- The procedure for reporting accidents, adverse events and incidents.
- The Data Protection Act (1998) and General Data Protection Requirement from 25 May 2018.
- Any other legislation relevant to your research. You must also ensure that you are aware of any emerging legislation relating to your research and make any changes to your study (which you will need to obtain ethical approval for) to comply with this.
- Obtaining any further ethical approval required from the organisation or country (if not carrying out research in the UK) where you will be carrying the research out. This includes other Higher Education Institutions if you intend to carry out any research involving their students, staff or premises. Please ensure that you send the FREP/SREP copies of this documentation if required, prior to starting your research.

- Any laws of the country where you are carrying the research and obtaining any other approvals or permissions that are required.
- Any professional codes of conduct relating to research or requirements from your funding body (please note that for externally funded research, where the funding has been obtained via Anglia Ruskin University, a Project Risk Assessment must have been carried out prior to starting the research).
- Completing a Risk Assessment (Health and Safety) if required and updating this annually or if any aspects of your study change which affect this.
- Notifying the FREP/SREP Secretary when your study has ended.

Please also note that your research may be subject to monitoring.

Should you have any queries, please do not hesitate to contact me. May I wish you the best of luck with your research.

Yours sincerely,

SREP Chair

Appendix 9

Ethics approval – ESPGR-08

Principal investigator:	Joao Guerra
Project supervisor:	Lee Smith
Project title:	The potential yield of external feedback via movement sonification in physiotherapy
SREP code:	ESPGR-08
Approval date	22/11/2019

Application decision: **Approve with revisions** under the terms of Anglia Ruskin University's Research Ethics Policy (Dated 8 September 2016, Version 1.7). Approval by SREP is subject to ratification by the FREP.

Changes to be made: These changes should be discussed and approved by your supervisor (all documents must be updated online) but do not need to be communicated to SREP, all changes must be made before data collection can start:

Ensure that the PIS sheets are updated in Box B3 to state that the data controller is AP not the student

All documents (PIS, Consent form, Debrief) given to participants, must be printed onto Anglia Ruskin University headed paper.

Any advert must contain the following statement:
The study has received ethics approval by the School Research Ethics Panel (SREP) and ratified by the Faculty Research Ethics Panel under the terms of Anglia Ruskin University's Policy and Code of Practice for the Conduct of Research with Human Participants

If you make changes to any aspect of your approved research, it is important that you discuss this with your supervisor as they can advise you on whether you need any additional ethical approval.

Ethical approval is given for a period of 1 year for undergraduates/masters students. If your research will extend beyond this period, it is your responsibility to apply for an extension before your approval expires.

It is your responsibility to ensure that you comply with Anglia Ruskin University's Research Ethics Policy and the Code of Practice for Applying for Ethical Approval at Anglia Ruskin University available at www.anglia.ac.uk/researchethics including the following.

- The procedure for submitting substantial amendments to the committee, should there be any changes to your research. You cannot implement these amendments until you have received approval from SREP for them.
- The procedure for reporting accidents, adverse events and incidents.
- The General Data Protection Requirement and Data Protection Act (2018).
- Any other legislation relevant to your research. You must also ensure that you are aware of any emerging legislation relating to your research and make any changes to your study (which you will need to obtain ethical approval for) to comply with this.
- Obtaining any further ethical approval required from the organisation or country (if not carrying out research in the UK) where you will be carrying the research out. This includes other Higher Education Institutions if you intend to carry out any research involving their students, staff or premises. Please ensure that you send the FREP/DREP copies of this documentation if required, prior to starting your research.
- Any laws of the country where you are carrying the research and obtaining any other approvals or permissions that are required.
- Any professional codes of conduct relating to research or requirements from your funding body (please note that for externally funded research, where the funding has been obtained via Anglia Ruskin University, a Project Risk Assessment must have been carried out prior to starting the research).
- Completing a Risk Assessment (Health and Safety) if required and updating this annually or if any aspects of your study change which affect this.
- Notifying the SREP Secretary when your study has ended.

Please also note that your research may be subject to monitoring.

Should you have any queries, please do not hesitate to contact me. May I wish you the best of luck with your research.

Yours sincerely,

Dan Gordon

SREP Chair

Date 30.9.2019

V1.5

Appendix 10

COREQ Consolidated checklist

COREQ (COnsolidated criteria for REporting Qualitative research) Checklist

A checklist of items that should be included in reports of qualitative research. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			
Personal characteristics			
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	205
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	205
Occupation	3	What was their occupation at the time of the study?	205
Gender	4	Was the researcher male or female?	205
Experience and training	5	What experience or training did the researcher have?	206
Relationship with participants			
Relationship established	6	Was a relationship established prior to study commencement?	206
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	206
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	206
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	206
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	206
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	206
Sample size	12	How many participants were in the study?	206
Non-participation	13	How many people refused to participate or dropped out? Reasons?	206
Setting			
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	206
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	206
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	207
Data collection			
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	206
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	206
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	206
Field notes	20	Were field notes made during and/or after the inter view or focus group?	206
Duration	21	What was the duration of the inter views or focus group?	206
Data saturation	22	Was data saturation discussed?	206
Transcripts returned	23	Were transcripts returned to participants for comment and/or	207

Topic	Item No.	Guide Questions/Description	Reported on Page No.
		correction?	
Domain 3: analysis and findings			
Data analysis			
Number of data coders	24	How many data coders coded the data?	207
Description of the coding tree	25	Did authors provide a description of the coding tree?	207-213
Derivation of themes	26	Were themes identified in advance or derived from the data?	207
Software	27	What software, if applicable, was used to manage the data?	NA
Participant checking	28	Did participants provide feedback on the findings?	No
Reporting			
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Yes
Data and findings consistent	30	Was there consistency between the data presented and the findings?	Yes
Clarity of major themes	31	Were major themes clearly presented in the findings?	Yes
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	Yes

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

Appendix 11

Focus Group

Transcript and questions

JG - My name is Joao Guerra and I am the moderator. This project was created as part of my PhD and I will start by thanking you for participating. I would now ask if you could introduce yourselves and say what your profession is, that would be excellent.

SA - My name is SA and I have been a physiotherapist for 19 years.

DC - DC, physiotherapist with 23 years of experience.

RM - RM, physiotherapist since 2003

MG - Hi, MG, physiotherapist and I graduated in 2007.

JG – Question 1 - Thank you and let's get started with our focus group. The first question is an open question so we can start with MG since he was the last one to introduce himself. The question is: What is your opinion about the ASSA prototype?

MG - Well, I think that the prototype, taking into account what is main function is, it achieves its objectives, it has its limitations in its potential applicability to reality, but for the objective of the study in which we took part, it looked to be very useful and it was easy to understand how to utilise it.

SA - I can say that I have a positive opinion of the prototype, even though I have some questions that we will discuss next, but above all I have the impression that the prototype was well received by the participants, and by those who were using it and that is always a good indicator that the prototype was important. So, I believe this acceptability indicator by those using it is important.

JG - So you noticed that during the study the participants also thought it was easy to use?

SA - Yes, I did. There were no constraints in the utilisation of the prototype and that is a favourable point because sometimes you can have something very reliable and that investigators think is very useful but if the users feel any constraint in utilising that, either because it takes too long to place correctly or something else, I believe that will then create some difficulties in the clinical practice.

DC - In general I too also liked the prototype, even though it is a prototype and before the final version is ready it can be improved. However, as a prototype and considering we're talking about an idea of what will be constructed, I think it is pretty good. I also felt that the participants that were using it were pleased as they are not used to being evaluated in that manner and thought it was much better than the traditional way that is used to evaluate strength. In general, this is it.

RM - I liked it, I liked it a lot. The concept of the prototype is an excellent idea and it can indeed bring added value if integrated in the physiotherapist's clinical practice. It's an interesting and useful concept. However, I agree with other's comments and there are some things, small and big, that can be improved in the equipment, but it was very useful and the practical utilisation was very well accepted by the participants. Even those that were not very experienced, they quickly adapted to what needed to be done and there were not major hurdles in understanding how the prototype worked, it was very intuitive.

JG - I would just ask this, taking into account your comments, comparatively to a standard strength test that you would perform in your clinical practice, do you think the prototype changes that dynamic or is that acceptable in order to use the prototype?

RM - It can change but it's mainly related to the equipment current hardware - by this I mean if small changes are made to the equipment, I don't think there would be any problem during the evaluation.

SA, DC, MG - Yes, we agree.

JG – Question 2 - That's great, let's move to the next question and feel free to share ideas too. The sound feedback feature was only used by two of you but what we want to investigate is how can the sound be understood by the physiotherapists and in which areas do you believe it

can advantageous? Consider two perspectives, the physiotherapist hearing the sound or the participants hearing what is happening and is that an advantage to the clinical practice in different settings of physiotherapy.

RM - In my opinion, I felt that the sound did not bring a lot of added value. Where I can imagine the sound can be useful is with people that do not have a lot of experience, who have recently graduated, or people that have a lack of perception - this does not necessarily depend on being a recent graduate but on our individual capacity of perceiving another's strength - then the sound effect can bring value because it helps the person to synchronise his manual perception with the audio feedback. In my case, I did not feel the sound made a big difference, but I don't know what effect that will have on the data.

SA - I actually have a different opinion than RM, I understand what you are saying, and the sound did not influence me greatly. However, I remember that the person who was doing it with me, M, the first time she used the prototype with the sound she said: Oh I realise I was resisting too hard in the beginning. So I remember that clearly and she also mentioned that this way she knew if she was applying too much resistance or not enough. But I am not sure this is very good, now that I am reflecting on it. Sound might create a false perception, so I am not sure about this. I also noticed that we did a lot of repetitions on the same day and this would not happen on a normal clinical setting - in real life we would do one test one day, then the next the day after or in three days' time. What I noticed was that the sound gave me an extra motivation – like a positive input, like when you increase a music volume for a difficult task, which I associated with running - so the sound in the prototype had that similar effect. The person who was performing the test with me mentioned the positive effect of the sound and that is important.

DC - My opinion is merely conceptual because I did not use the prototype with the sound. Theoretically the sound can bring a lot of value if there is the option to turn it on or off because in some situations the sound can be extremely good, and in others it can influence the results. If the person follows the sound to achieve the speed they want, this can somehow influence the results obtained - which is not necessarily bad but I am not sure that when evaluating maximum strength this is good. So when SA mentioned that M had realised she was putting too much resistance in the beginning when she used the prototype with the sound, so this changed her evaluation.

318

SA - Let me clarify, what M said was that when she was placing the prototype, and even before asking the person to initiate the movement, she realised she was applying too much pressure already as the sound had started to play.

RM - To clarify, regarding the comment from DC, the sound does not regulate the speed, it only regulates the strength. When the person applies more strength, the sound's frequency gets higher, when the strength diminishes, the sound's frequency gets lower.

JG – So, just so everyone understands, the prototype can have sound, or not. What we were trying to investigate was the differences in sound mapping. We were going to test one type of sound mapping, directly correlated to the strength and another correlated to the speed. There is some research that says the speed at which the test is performed can affect the reliability, specially at an inter observer level. Evaluator X applies a certain speed and achieves a maximum strength, but evaluator Y might apply a different speed and the maximum torque will be different. So we wanted to know if the average speed would affect the reliability, but we did not perform that test. What RM and SA did was the evaluation of muscle strength, so the higher the force applied, the higher the sound pitch.

MG - So there is no sound signal for the beginning and end of the test?

JG - No, because since the movement range changes with each individual taking the test, it's difficult to define it that way.

MG - That could potentially be beneficial, because patients have the tendency to apply maximum strength, and almost kick, so when they hear that sound initially it could be beneficial.

SA - What's interesting to notice is that continued to happen - so in M's case because the sound showed she was applying strength in the beginning, she stopped applying that initial force. She even said: "This is more interesting as I can now see I was applying too much resistance even before the movement started". So it was the opposite in that case.

DC - I have one question, are you supposed to have a minimum and maximum value for the sound and would the frequency depend on that?

JG - In sonification, the objective is to be able to attribute the parameters that one considers to be more relevant. With sound, there are so many parameters and possibilities that it is impossible to explore them all.

DC – Let me rephrase, maybe I did not explain myself very well. What I meant is, if I am evaluating the force associated with the wrist flexor or the finger, compared with the quadriceps, the magnitude of the force are completely different. My question is, does the prototype distinguish the different levels of minimum and maximum strength?

JG - The answer is yes, I created one mapping for the knee and a different one for the elbow, precisely because of what you mentioned.

DC - Ok, great, then I think that is one of the greatest advantages of this prototype compared to what is available in the market today, at least that I know of. More than being useful purely for the evaluation of strength, it's useful for the daily work with the patient. In my opinion this prototype can definitely make the strength evaluation more reliable, but the major advantage is for when a physiotherapist wants to work with a patient within a certain pre-defined value of force production within a specific range of motion. At the moment, with this prototype I can use the sound frequency to say that I am working at 80% RM and it's easier to make the patient achieve its goals then with a manual testing alone. So more than for the sole evaluation of strength, for the daily work with the patient this can be truly revolutionary and advantageous for the patient.

JG - So taking DC point of view, I'd like to ask you to think in which area of physiotherapy do you imagine the sound feedback could be used. We know, for instance, that sound feedback is used with patients with Alzheimer's, so can you see any area in which this could bring advantages for strength training?

RM - I suppose it could be useful in the more acute situations – as soon as the patient is admitted to the hospital, I am referring this with intensive-care in mind, especially at the initial stages of the strength recovery it can be particularly important. But I can also see advantages in any other phase within the hospital, where function gain is at its earliest stages.

SA - Can you explain why?

RM - So with some changes that would enhance the prototype and make it more accepted by the healthcare professionals, the sound feedback allows you to have a more rigorous control of the grade of strength applied to the manual resistance for each individual. Instead of applying the same generalist treatment to every patient, you can now adapt the strength training and take into account inter-subject differences.

SA - The is true but that issue can also be applied to other areas.

RM - Yes, but you now have an equipment can reliably quantify the measures

SA - I understand but I am struggling to say if this could be better applied in a certain area versus another. Naturally what comes to my mind are the areas in which I work and in those I can visualise the use of the prototype, like in intensive care. But I can also see it being used in other areas, as it is quite versatile.

RM - If we now talk about physical exercise, just like DC was saying, I agree with him, with some changes the prototype can be an incredible improvement in areas like with community-based physiotherapy, visiting the patient where there are not a lot of available resources.

DC - Not only in a home visit. There are other handheld dynamometers that allow you to measure strength in a reliable way and that allow for a daily development of strength. The biggest advantage of this one, is that it is so much more than just an HHD, as it allows you to evaluate speed, which in some situations is highly advantageous. It also allows to evaluate dynamically and gather information the arc of the moment and not just one position. The prototype can do more than the traditional manual strength testing because it can do a dynamic evaluation - I am not constrained to one standard position, I can evaluate the strength applied in the execution of a certain movement, and not just the muscle strength in one task and that is a fantastic advantage.

I don't think in terms of "Where can I use the prototype" as it can be used in multiple different areas. We need to consider "What is the prototype's purpose" and so the prototype will be reliable and fit to be used for that it was designed for.

The prototype has the advantage that it evaluates strength exactly as a dynamometer would, but it also adds other elements that the dynamometer can't - for instance it evaluates speed and this then becomes not a question of where it can be used, but how. The prototype allows to evaluate strength in a traditional way, as well as in a dynamic way and using the whole amplitude of movement. It also allows to evaluate specific situations, such as in the case of a stroke patient - how is the strength changing with the speed of execution - which gives me a lot of information on the quality of movements as simple as taking your hand to your mouth. So specific tasks can be evaluated that I wouldn't be able to do with a traditional dynamometer.

RM - I meant the intensive care examples as it meant the prototype could be used at the initial stages of the intervention – for initial strength gain. There are limitations to

the prototype though, specifically in the functional area as I don't see this being constantly used by a physiotherapist when applying manual strength testing in a patient, as the patient will eventually evolve. So I consider the prototype will be very good in the initial stages of strength recovery and the training properties DC mentioned will bring value - the fact that you can find the maximum strength a patient can apply and then make sure he applies that during the movement arc. Similarly, to the isokinetic, we can use the prototype to change the force applied during the movement, depending on the size of the segments, obviously. One of the issues we see sometimes is that people start the training by applying a certain load without really knowing why they are doing with that load. With a prototype that measures the force at an initial stage and that then allows you to apply that same strength along the range of motion, it is possible to make sure that during the movement we are always applying the ideal force for that muscle and that specific angular position, and that is the value of the prototype. Not only assessment but also muscle training tool

DC - The big advantage of the prototype, when compared with the isokinetic, is that the isokinetic is too static. With the prototype you can explore functional tasks that you couldn't with the isokinetic with specific tasks which an isokinetic couldn´t - example of taking the hand to the mouth. There might be some disadvantages when compared to other equipment's with higher reliability rates like the gold standard, but there are still advantages over the gold standard as it can evaluate things the isokinetic dynamometer can't.

For those that work in sport area, if I consider of an athlete with a key technical gesture, if I can break that gesture up into more than one functional part, that can be evaluated with the prototype. Maybe a rugby tackle, the individual parts of the tackle, for instance, one might be able to be divide that into smaller parts for it to be assessed. If I can understand the differences between the athlete with the perfect tackle technique and other athletes, I can the more easily train the other athletes to achieve the best possible tackle technique.

JG - MG, would you like to add anything to this?

MG - I agree with what has been said - where I use the value it can bring is not so much in what RM said like the intensive care as there can be limitations to the prototype

there, since patients can be intubated, with intravenous access, or there might be other electronic equipment that could interfere with this one, I don't know.

In high-performance sports, I think it is very useful to understand at what angular position is the maximum strength peak achieved - there are a lot of situations in which it might make a lot of sense to know where can an individual reach its maximum strength or if he should increase this strength at a different angular moment and that is a good example of where the prototype can be used in high-performance area. It can tell for instance that an individual has an excellent strength capacity at 90 degrees but then at 110 degrees it drops dramatically for what would be expected for that specific task he should be performing. In that case, it really has more potential than others.

JG – Question 3 - This is related to the next question - do you picture yourself using the prototype in your workplace and if not, where would the problems be - buying the equipment, its maintenance, the patients' acceptability.

MG - In one of the areas where it could be interesting to use the prototype, but perhaps the second or third iteration, is in the water, depending on how water-tight it is. I don't know anything like this that could be used in hydrotherapy. The portability of the prototype is another factor - I don't know if you're planning to make any alterations - but maybe there could be a softer handle to hold the prototype. The fact that it has the wires, and its weight - will these be exactly as in the prototype?

JG - My question was more related to the concept of the prototype, rather than its characteristics - do you think it would be acceptable to use at a teaching level for instance, with students, or in a care-home, in a sports club, do you consider people would use it?

MG - Anything that is related with research or high-performance, this prototype would be a perfect fit. For the clinical practice of a physiotherapist that has to move from one place to another, I don't see it being used because even something as simple as clinical records is hard for some physiotherapists, so it would be the same with using these type of equipment.

DC - I disagree.

SA - I also disagree. I believe the prototype will be easily accepted, it's simple, educational. Even with the participants it was well received straight away. There are some problems as MG mentioned like the wires and the weight. But that's more the characteristics, not the concept.

DC - In terms of concept, I think it is fantastic and it will be very well received. Because if someone tries to sell me something that can give me specific data of what I am trying to evaluate and even other parameters that maybe I didn't even consider to evaluate in my clinical practice, also with an immediate registration of the data that can be accessed immediately and at any time, then that's something I'd want. And, more importantly, it does not change greatly what I already do on my clinical practice. Because If someone came to me in my clinical practice trying to sell an isokinetic, it would require me to learn how to use it, I would have to use it in a specific location, would have to do the setup, etc. But with this prototype I simply have to use something between my hand and the patient, instead of using only my hand. So, with regards to the concept, the fact that I would only have to use something to hold between my hand and the patient and that can give me the right data, means it will be accepted widely. There are certain constraints, like the price, and things that can be improved in the prototype - namely its functions and applicability. Taking into account those improvements and the price, there will always be the question - why should I pay a certain value to use manual muscle tests with the HHD, when can use it as I use now and still get results, even if worse ones.

MG - The concept I was describing included the computer, and what is needed now to use the prototype. But I see this as a great opportunity for many areas.

SA - DC was saying that the question of "why should I buy this if I can simply use my hand" will arise. We need to contemplate other marketplaces, not just the Portuguese. If you consider the American market, if you need to present results of your evaluation, then people would quickly realise this can bring value. Portugal's reality is changing - when we say the patient will need 12 physiotherapy sessions, there's the question, why not 10 or 6, etc. there will be the need to show that what you are doing truly if working.

MG - For the insurance business, for instance, it would be great.

RM - The concept and the device are good. But not matter how useful it is, or how interesting it is and how important it is in terms of advancing our scientific knowledge and clinical practice, at the end of the day what matters is the cost. When I was working

at Kyneticos, people would get very excited when we showed them the product, but as soon as the price was mentioned they would say they were not interested.

SA - I think the issue is the price and the fact that you feel obliged to show results, these are connected. If there were more pressure to demonstrate results, people would consider the price differently.

RM - At the end of the day it goes like this: what my client wants to see is this. So how can I maximise my investment in this equipment. From the moment that people realise that a certain equipment costs x but will bring in value of x plus something, that's fantastic, people will buy it. If people can't see this, they won't purchase the equipment, put simply. What DC said is important, if the equipment does not add bring any significant value to the clinical practice, people will not buy it.

SA - I understand that.

RM - The other thing is maintenance. This is a unique prototype, and that is valuable. So it's different if you purchase a product that is yours and you own, or a product that you then have to pay an annual fee to maintain, at that point it becomes a service. I believe this prototype could be seen as a service.

JG – Question 4 - These are very interesting answers and they answer question 4, which is about what would make physiotherapists not want to use the prototype in their clinical practice. RM has already answered this. Going back to the other participants, what aspects do you see as being an obstacle to the acceptance of the prototype, be it in the financial aspect, maintenance, in the different settings like hospitals, clinics, in the community, etc.

SA – To me, one factor is inertia. That will always be a problem. At the beginning people will find it interesting but after some time they will go back to what they do every day. They will keep doing what they did in the past and that can be a big limitation in our profession, and perhaps even in others, and needs to be taken into account as this can impede people from using it. Another thing is that the prototype has to be easier to handle - the device´s portability is important.

JG - Going back to SA's comment on the inertia, can you expand on that or maybe suggest strategies that could help in the dissemination of the prototype or in introducing it to the clinical setting?

SA - I would really like to have more suggestions. But I Portugal it's taken us 10 years to convince people of the importance of clinical registers, and this was only achieved when it became mandatory in hospitals. This is the reality, so I don't know how to answer your question. Maybe marketing campaigns could help. Another thing is if we can convince some of the strategic key people that usually sell these ideas in physiotherapy to use it, that would be important. I also think it would be really valuable to present this at the WCPT congress next year, in Dubai. I had the experience of attending last year's congress and it was light-years from what we currently do. So maybe you need to try to sell this at a place that can launch your work and prototype. I believe the Portuguese market is too small for this, and I really don't have experience on the international market to give you an answer on that.

MG - It depends on the perspective, there can be a lot of limitations if there is a lack of interest, and not a lot of limitations if people are really interested. The first thing, I agree, is the inertia. Another thing is the unknown factor, that can be relevant too, the willingness to try the prototype - some people will say they don't want to learn about the prototype because it will be seen as extra work when performing muscle testing, and they can still perform the test without it, grading it from 0 to 5 and doing it quite quickly. The time it takes to utilise it, can also be a factor, as with the prototype the physiotherapist would spend more time when compared to the traditional muscle testing. The device´s mobility can also be an issue, since it would mean I would have to carry one more thing. There are more factors, related to the possible difficulty in the utilisation of the prototype, the weaknesses of the equipment - like people not feeling very comfortable during the evaluation because they do not have good manual contact with the equipment. I believe it depends a lot on the initial interest in utilising the prototype.

In terms of using the device for clinical records, particularly from an insurance companies perspective, if the interest is there for physiotherapists to use this, then a lot of these obstacles will disappear.

RM - This will more easily get accepted into the market if it comes from the top, rather than from the bottom - if insurance companies start saying that this is the best practice, some clients will accept that and others won't.

SA - Before that you need to have 2 or 3 key people writing articles about this, so there is still some way to go.

RM - Yes, true but there a lot of potential here.

JG - DC, would you like to comment on this, to remind you, we're looking at what would make people not want to use the prototype in their clinical practice.

DC - I imagine people would use this, no doubt. Unlike what SA said, I believe you should start by developing a case study in Portugal because if you can sell this in Portugal, you can sell it anywhere in the world. Almost everything else has already been mentioned.

With regards to buying this, there will always people around the world that will like the concept. You were able to convince us and the students easily so a lot of people will also be easily convinced to purchase the equipment. The dissemination of the prototype will happen automatically.

What can prevent people from buying the equipment? Quality of the product, some weaknesses but I know we will discuss that later. The concept and the idea of the prototype was well created, and it is a winning one, it will sell. Now you need to apply that to reality and make the prototype 100% functional, of extremely high quality, even if it makes it more expensive, it needs to be high quality. We're talking here about a piece of equipment that is portable and will be carried from one place to another, it will have to be quite resistant to falls as that might happen often. What happened during our data collection was that sometimes we had to stop and restart it, and that can't happen. RM was talking about the support and maintenance so people won't purchase something that will mean a change to their clinical practice but that then has problems every week like a bug or something and needs to be repaired - that can't happen. And also, if the equipment is really good, but is not very durable, that won't work either. We talked about this before, with regards to the wires, the prototype will need some kind of battery or perhaps even two, so you can use one while the other is charging because you don't want to make a patient wait for the equipment to be working. It's about trying to anticipate these kinds of problems.

JG – Question 5 - This is great feedback. Moving to the last questions, let's try to summarise what you said and think about what did you enjoy the most about the prototype, what did you enjoy the least and what do you imagine could be added or removed.

SA - What I liked the most was the concept and how easy it was to understand what needed to be done. It was quite intuitive and easy to use. To me the main constraint was the wire and the way to hold the equipment, this needs to be improved. Without the wire, the other concern would be the autonomy because if I need to use it but forgot to charge, then there needs to be a way to quickly overcome that otherwise I might not feel inclined to use it again. As for the audio feedback it would be important if it worked with any headset, not just some specific ones. So it needs to be adaptable in that way.

JG - Just to clarify something, when you mentioned the concept, did you mean the fact that it can evaluate force like a normal handheld dynamometer and the fact that it can capture movement amplitude?

SA - Yes, I meant all of that and all of the potential, not just in clinical practice but also in research, as it can play a role in that area.

RM - To summarise, the concept is very good. One of the most positive factors is that it is quite intuitive to use, and that is important for clinicians' acceptance, as people do not have time to learn new things - the more intuitive it is, the better. The prototype has a lot of development potential because of the quantity and type of hardware that it uses, what we are using as a prototype is about a third of what it can truly be achieved with all the hardware that's included. And that is what I like the most, the fact that it can become more versatile and have more applications. What I enjoyed the least was the ergonomics, the weight and the fact that it is not wireless. It could be interesting to develop a wireless way to communicate not just with the laptop collecting the data but also other systems. Let's say I already have a data collection system in my clinic, I'd like to be able to integrate this prototype with what I already.

JG - RM, when you say systems, you mean software already used in physiotherapy?

RM - I meant, software commonly used in physiotherapy, but also smartphones or smartwatches. If you can develop this in such a way that it can integrated with a

smartphone application, then you don't have to use with such large, closed equipment (like a computer) which will help with the transportability. If it can be integrated with these kinds of gear or other data collection software, it would be good. When we are discussing the cost, one thing that can have an impact on cost is this - if I already have a system in my practice that collects data and does the register, I am only lacking an equipment to collect data on force in a rigorous way. Then this prototype gives me this and if I can integrate it with what I already own, it's another advantage for the clinicians who buy it. Even if the price is a bit higher, it would be perfect, since I already have equipment that can be integrated with the prototype.

SA - It also needs to work for people who don't have other systems.

RM – Yes, if you don´t have anything, people that have an Android phone, then they can just download the application from the Appstore.

MG - As others said, the most innovative aspect is the concept. We live in a digital time, and it's important to have the potential of these type of digital equipment. Something valuable would be to develop software that it's easy to use and that anyone could easily pick up on, even if you have no expertise with this technology. I had also thought about what RM said about it being compatible with different applications. One example is the kind of platforms that hospitals like CUF or Hospital da Luz (hospitals in Lisbon) have, which means physiotherapists could register the data in those platforms and doctors would have access to this data live when they had appointments with the patients. So, there would be a way of communicating between the doctors and the physiotherapists.

The prototype's portability, and autonomy are something that needs to be improved, as that is very important these days. Another thing that can be improved is its ergonomics. Also, the part of the prototype that contacts directly with the patient's body could have some sort of cushioning or should be improved in some way.

JG - Do you mean to suggest that the shape of that part could be adjusted to the specific limb? Because there was already some sort of cushioning in the prototype.

RM - The foam was perhaps too hard and did not bend or adapt to the patient's segment.

MG - It could be something more bendable, that could adjust to the body part it was being used on. If for instance it was being used on the anterior lower part of the lower

leg, some people will have specific shape for which the prototype can become uncomfortable.

RM - So maybe something like a gel type of material.

DC - Or instead of being that high-density foam, maybe a low density one.

MG - Something that could be easily deformed for 1cm or so.

SA - We also need to consider that these days, whatever type of material has contact with the patient, will need to be washable and easy to disinfect.

MG - I will also add that the prototype should be watertight, maybe for its use in hydrotherapy as well. Not necessarily so that it can go under water, but that can resist splashes, at the least. So if there is an accident and someone spills something on it, it can at least not be damaged by that.

JG - Ok, so being impermeable.

DC - Ok, so answering the question. As everyone said the biggest value of the prototype, if it gets further developed, is its potential. With certain modifications, many of which we already discussed here today, more than a winning concept, you have a winning product.

In terms of questions to raise, I would say the weight is still a problem as it is a heavy item. When it comes to ergonomics, you need a handle that can be easily adapted to the physiotherapist's hand, like a ball handle or another type of handle. For the part of the prototype that gets into contact with the patient, it should be something removable so it can be adapted to more than one area. You don't need a lot of different options, maybe one plain, one semi-round and one that would be rounder would be enough. If you have these adaptions it's better in terms of safety. As we saw with the tibia, there was the risk that the prototype would slip away if the patients pushed too hard, so this would help in that safety element. That the high-density foam was a good idea and that it is better that the gel or low-density foam option as these give a false sense of comfort. Yes, it might be more comfortable at the beginning of the movement but as the movement progresses, the foam will be squished and not have much effect. So, the high density is a better option, even if it is more uncomfortable at the beginning, at least it will be better during the rest of the movement. Maybe an alternative would be a mixed type of foam, with a low- and high-density section.

Also, when it comes to the portability of the prototype, it should definitely not have the wires. If it was wireless then you can start considering some kind of box, and even a kind of power bank so you can charge the batteries without needing to plug them to the main socket. And you should have extra batteries too, so people can buy the equipment with just one battery or multiple batteries. This prototype should "grow" with the person who buys it, so unlike other kit where people have to buy the whole package even if they don't need everything initially. So, you can add batteries and power banks after you purchase the basic prototype.

In terms of the prototype being easy to use, I think it's worth noting that when you first designed the prototype, you could not anticipate everything. You thus need to keep the prototype in such a way that it's easy to change and develop further.

I don't agree that it needs to be very watertight but I do agree that it needs to be shock resistant. Maybe adding a strong type of covering to avoid that damage in case of a fall. In terms of needing a laptop, this needs to be changed too. The laptop software you developed could then be used in an app, or on laptops. The software needs to give the possibility to export the data, not only to other types of software but also to clinical databases, such. As the CUF example. It should be possible to easily send the graphs or data by email, or upload to some sort of platform, and ideally all done from an app. This means physiotherapists don't need the carry a laptop but can use their smartphone and see the results in graphical form straight away. We did not have that; we only saw the numeric values of the data and it was hard to do the evaluation and look at the numbers. I believe it would be important to be able to have a way of displaying the information during the evaluation, perhaps with a very simple graph on a smartphone or even on the prototype itself, some dynamical graphical element that would change as the evaluation progressed. So, you then have the audio feedback but also instantaneous visual feedback, that would be a great way to improve the prototype. And finally, you need to invest in the equipment's quality and durability. If some of the changes are made and the equipment is durable and of high quality, you don't need to worry about divulgation or trying to sell it at congresses. I know it might be difficult to do this, but if you could somehow "professionalise" the development of this concept into a final, high quality product, then I have no doubt, it will be very easy to sell these. You don't have to sell the concept or idea; you need to work on what you have and develop it into its final shape.

JG - I will now summarise what was mention during these last three questions.

What you enjoyed the most was the concept; the fact that it can provide information in different areas like force, velocity and angle; the fact that it is intuitive; the fact that it enables to capture data for clinical records and its potential.

All – Agree.

JG - What you enjoyed the least, everyone mentions the weight, ergonomics, the wires, the prototype's portability and its way of adapting to the specific patient segment.

All – Agree.

JG - Lastly, in terms of the characteristics that can be added or removed, everyone mentioned the autonomy, so adding a battery, removing the wires, adding a way to listen to the sound with any type of equipment like headphones, not using a laptop if possible, to have an automatic output, integration with other equipment or systems, adding a visual feedback, being water-tight, making sure the contact surface is cleanable, purchase options to include customisation and design that can be easily adapted. Anything you would like to add?

All – Agree.

JG – If there is nothing else to add, we can end this focus group now. Thank you all.

Appendix 12

Ethics approval from HSHPRC

Comissão de Ética da ESSCVP

Parecer Nº16 /2019

SOLICITANTE: Joao Guerra

ASSUNTO: pedido de parecer para o projeto - **O potencial rendimento de feedback externo via sonificação do movimento humano em fisioterapia**

AUTORES DO PROJETO:
Joao Guerra
Dr. Lee Smith
Dr. Genevieve Williams
Dr. Domenico Vicinanza

FORMA DE COMUNICAÇÃO: e-mail

A Comissão de Ética da ESSCVP, analisou a documentação do projeto de investigação intitulado - **O potencial rendimento de feedback externo via sonificação do movimento humano em fisioterapia**, e o parecer é favorável à realização do estudo.

Lisboa, 18 de novembro de 2019

P'la Comissão de Ética, a Presidente

Ana Paula Nunes