Inter-rater reliability using a modified Balance Error Scoring System

Lily Purdon

A thesis submitted in partial fulfilment of the degree Masters of Applied Science
at Otago Polytechnic, Dunedin, New Zealand

Abstract

Falling is a common occurrence among older adults with over 1/3 of older adults falling per year. There are a range of balance tests used to indicate an older adults risk of falling, many of which are attention invested and arguably do not apply to real-world settings. In real-world situations an individual's primary focus cannot always be fixed on a balance task. This study used divested attention by combining the balance error scoring system (BESS), an established balance test, with the controlled oral word association (COWA) test.

The primary aim of this research was to examine the inter-rater reliability of a modified BESS (mBESS) test. Secondary aims included determining if there was a relationship between the BESS and COWA scores, and whether this test was an appropriate level of difficulty for older adults.

The mBESS test was administered to four older adults, each completing three recorded sessions involving three trials of each balance stance. These recordings were then submitted to four volunteer raters who scored the tests independently using the prescribed BESS tools provided. Inter-rater reliability was assessed using Pearson's correlation coefficient.

The average total inter-rater reliability was 0.97, with the lowest reliability observed when scoring the one-legged stance (0.83). As expected, the variability between raters increased with the difficulty of the balance stance task. The mBESS test therefore offers excellent inter-rater reliability. Due to Covid-19 and the nationwide lockdown interfering with data collection, the secondary aims could not be addressed.

The mBESS test has excellent inter-rater reliability (0.97) and the test could be suitable for future research regarding falls prevention to help assess falls risk with more authenticity.

Acknowledgements

I would like to begin by acknowledging all the beautiful people who have helped me throughout my thesis process.

Firstly, Phil, my amazing supervisor who has inspired me with his knowledge and critical thinking. Both your and Richard's ongoing support within my thesis and other projects has made this year all the more positive.

To my Mum, Dad and family, for putting up with me through my stressed moods, as well as hearing and reading my thesis over and over. I couldn't thank you enough for providing me with all the love and support, especially when I needed it most.

Lastly, to Jess and Rachelle (the bits)! Thank you for being open ears and always knowing how to pick me up during my stressful days. Your happiness and humour will always brighten my day, love you lots.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Due to the growing concern around elderly falls and the potential harm it could cause, this project explored the possibility of a new test that could accurately evaluate an older adult's fall risk. This research was designed to examine the reliability of the combination of two existing tests, the Balance Error Scoring System (BESS) and the Controlled Oral Word Association (COWA) test. The combined test was referred to in this project as a modified BESS test (mBESS). This dissertation reviews existing research on balance testing to help provide an understanding of why the mBESS was being explored, the use of this test, and why these two tests were chosen.

Covid-19

Due to the nationwide Covid-19 lockdown this research was modified from the initial proposal. The New Zealand community lockdown occurred two weeks into testing and therefore forced data collection to cease. Subsequently the research design had to be altered to fit the programme time frame. The research therefore changed from being a reliability study to become an inter-rater reliability study. The specific objectives were to determine the reliability of the modified BESS, compare individuals' BESS and COWA scores and determine the ease of administration and difficulty performing the test. The revised objective of this research were changed to 'determine the inter-rater reliability of scoring the mBESS test'. The literature review subsequently extended to include reliability testing, with a section on inter-rater reliability and learning effects to accommodate the modified aims of this research.

Background

Older adults and falls prevention

Old age or becoming an older adult is defined by the World Health Organisation as anyone over 65 years of age.[1] The older adult population is expected to increase, and it is predicted that by 2051 there will be 1.14 million older adults in New

Zealand, making up 25% of the population.[2] With ageing comes psychological and physical deterioration, both of which can be a factor contributing to falling.[3–5] Falls are the second leading cause of accidental injury or death worldwide, with 193,974 falls claims from older New Zealanders in 2017 [1,6]; the total cost of these falls was estimated to be over $267 million.[6]

There are many reasons why falls become more common with old age are often categorised into intrinsic or extrinsic factors. Intrinsic factors include any psychological, physical or physiological impairments, whereas extrinsic factors are environmental issues including, lighting, flooring, polypharmacy and footwear.[7–11] The strongest predictor of whether someone is likely to be at risk of falling is their history of falls.[8–11]

Falling can cause many serious issues including fractures or tissue damage, functional limitations, decrease in independence and in the worst case, death.[11,12] Not only can this be a burden on the individual, but also a huge cost to society.[6] Falling is financially covered by ACC, which is funded by the government, investors and New Zealand businesses.

If an older adult fell and returned to independent living they would be assigned a needs assessor to visit their house and assess what they could do to 'safe-proof' their home. This means physically altering their environment, through household, behavioural or medicinal changes in order to lessen the potential harm that could be caused by the individual falling again.[13] Although this service could help reduce the severity of the outcomes from falling it does not help address the root of the problem. Another strategy commonly used is exercise. There are many falls prevention exercise classes around New Zealand targeted at older adults. With the 'Live Stronger for Longer' initiative, partnerships had been created to help advertise 900+ exercise options nationwide.[14] The majority of these classes focus on improving cardiovascular performance, muscle strength, co-ordination and balance.[14] Although both the needs assessor and exercise interventions could decrease the harm of falling, only exercise addresses the actual problem of preventing the individuals physical ability to avoid falls compared to a needs assessor changing the physical environment.

Current balance assessments

Many exercise classes evaluate their participants' balance ability and risk of falling using various tests. Common tests used nationally and globally include the Timed Up & Go (TUG), the Four Staged Balance Test (FSBT), 30 Second Chair Stand (30 CST) and the Balance Error Scoring System (BESS).[15,16]

Although TUG was commonly used, there are multiple studies that believe it offers an inadequate prediction of an individual's risk of falling.[17–20] The FSBT, 30 CST, and both BESS have all shown strong reliability.[21–24] The FSBT, the BESS, also had intra- and inter-rater reliability evaluated, with all three tests showing strong intra- and inter-rater reliability.[24]

Although all of these tests measured an aspect of balance, whether it was mobility, strength, muscular endurance or flexibility, they were all conducted in a controlled setting. All tests were performed without distractions and therefore arguably did not represent the real-world balance demands.[26] This therefore raises the question as to whether they are valid tests for measure an individual's fall risk. Falls can commonly happen due to multitasking or becoming distracted and this ideally should be translated into testing to create a more valid test.[7,8,10,11] Current tests have low external validity and therefore we question whether these tests can provide sufficient insight into an individual's risk of falling.

Divested Attention training

Divested attention training, also known as dual-task training is the idea of focusing on multiple tasks at one time.[27] Divested attention training is used to create interference between two tasks being performed simultaneously.[28] This forces the participants' attentional demand to be spread between multiple tasks and therefore can alter the quality of the performance of each task.[29]

Divested attention training has shown to be successful among participants that are rehabilitating, recovering from lower limb amputation or losing functional movement due to age.[28–31] Research conducted among the older adult population showed the benefits of divested attention training for the gait cycle and posture. Divested attention training could therefore be ideal for older adults as falls often occur due to distractions and multitasking.[28] With divested attention training, older adults' task performance will be compromised and result in a decrease in performance of either one or both of the tasks, this is due to older adults ability to multi-task as they age.[32,33]

Inter-rater reliability

Inter-rater reliability is an important aspect to explore during test development. Having strong inter-rater reliability means that different testers can score the same test and record the same results.[34] The inter-rater reliability of the BESS was shown by Kleffelgård et al.[24] to be excellent with an interclass correlation coefficient (ICC) of 0.81.

Significance of the study

Falls are the second leading cause of injury or death worldwide and prevention strategies are important and require evaluation.[18] This research project combined a respected balance test with cognitive distractions and examined the inter-rater reliability of this mBESS. Currently there is no known research that has explored divested attention training and older adults' balance ability. This research avenue needed to be explored due to its success with other populations.[28–31] This research could lead to new strategies for falls prevention among elderly and therefore help to reduce the harm and number of falls occurring.

Research question

The aim of this study was to investigate the inter-rater reliability of the modified BESS test.

Chapter 2: Literature Review

Chapter overview

This literature review explores aspects surrounding the inter-rater reliability of the mBESS test. Firstly, it details why there is a need for this test by interpreting the literature about older adult health and falls prevention. This literature review then discusses the current balance assessments in use and their validity and reliability. The concept of dual-task training is discussed in regard to how this could be beneficial for older adults. To conclude this section, the current literature on inter-rater reliability testing and learning effect was reviewed.

Older Adult Health and Falls Prevention

Becoming an older adult is defined as anyone over 65 years old.[5] Ageing is a part of life which is accompanied by psychological and physical deterioration and an increased risk of chronic illnesses.[3–5] It is expected that the world will have a substantial increase in the number of over 60 year olds.[3,5] Statistics New Zealand[35] predicted the population of over 60 year olds will reach 1.14 million by 2051; this would equate to one in four New Zealanders.[2,3,35]

As well as an increase in chronic illnesses, some of the psychological and physical changes that can occur with ageing include slowing of reflexes, deterioration of eyesight, decline in mobility and decreased muscle strength.[7–11] Although many of these physical changes occur naturally when ageing, some may be prevented with early detection and treatment.[36–38] Enduring one or many of these changes can make the simplest of human movements difficult, including maintaining good posture or walking. It is therefore no surprise that these changes can lead to falls among older adults.[27]

Falling is defined as an unexpected event in which the individual comes to rest on the ground or lower level.[3,39] The WHO[1] believes falls are an important global concern as they can lead to injury or fatality.[40] The risk and seriousness of falling increases with age among older adults due to some of the changes listed above.[3,41] It is believed that 1/3 of community dwellers and over 1/2 of rest homes residents fall annually.[11] Death due to fall-related injuries contributes to 23% of deaths in over 65s and 34% in over 85s.[11]

The reasons why older adults fall can be categorised as having either extrinsic or intrinsic influences. Intrinsic factors include any psychological, physical or physiological impairments. The most common intrinsic risk factors include balance disorders, previous injuries, reduced mobility, decreased muscle strength, visual decline, cognitive impairment, dizziness, interaction of medications and existing illnesses.[7–11] Extrinsic factors include floor rugs, slippery or uneven surfaces, poor lighting or footwear.[9–11]

The strongest indication that an individual is at risk of falling is a history of falls. If an individual has fallen within the last year, they are at a higher risk of falling again.[7–11] When older adults fall it can result in fractures and soft tissue injuries. Long term affects include functional limitations, a decrease in confidence, loss of independence and ongoing costs for rehabilitation or help.[11,12]

Interventions to reduce falls include reducing environmental hazards, reassessing medications, modifying individual behaviours and exercise programmes.[11] If an older adult falls and returns to independent living it is common they will be assigned a needs assessor to facilitate those interventions, with the exception of exercise.[13] One issue with this approach is that exercise is the only intervention aiming to resolve the root of the problem. Current falls prevention exercise programmes tend to focus on balance, strength, mobility and aerobic capacity. Many studies have shown exercise programmes with compliant participants reduce the risk of falls occurring.[7,10,11,42,43] It has been noted that exercise programmes are the only effective intervention at reducing the number of individuals that fall, and reducing the rate of falls occurring among fallers. All other interventions tend to reduce only one of these variables.[10,11,44]

Older adult exercise programmes have been altered and refined to specifically help them perform everyday activities safely.[36–38] Although many exercise programmes for older adults can include a mixture of balance, strength, mobility and aerobic training it is hard to assess their effectiveness.[7] Many studies have used falls history to determine whether interventions are successful, but very few use a reliable test of balance ability.[7,10,11,42,43]

Balance Assessments and their Validity

Balance is defined as the ability to remain in a position without falling or losing control.[3,45] Balance ability decreases with age due to a greater influence of intrinsic factors that can impair their psychological, physical or physiological capabilities.[7–11] Balance is an important factor for older adults as falling is the leading cause of injury and can result in significant personal, social and economic burdens.[18] Within New Zealand a range of exercise programmes for older adults are offered, with the Live Stronger For Longer partnerships offering over 900 options.[14] Several common balance tests are used in New Zealand, and also globally, to help assess older adults' balance. These include TUG, 30 CST, FSBT and the BESS test.[15,16]

The Timed Up and Go

TUG involves the participant sitting in a chair, before standing up, walking around a cone placed 5 metres away, and then returning back to sitting. The risk of falling is assessed based on the time it takes the participant to complete that task.[46] Although TUG reflects the participants' strength and mobility, it does not address many other factors that contribute to falls.[18] TUG is recommended by organisations such as the American Geriatric Society, British Geriatric Society and National Institute of Clinical Evidence.[18,19] Evidence indicates that the TUG is inadequate at accurately predicting someone's risk of falling, yet it continues to be used and recommended by influential organisations.[17–20]

A key issue with TUG is the low sensitivity and specificity of the test.[18] Many studies use different cut-off times depending on their participants and this therefore makes it harder to adopt for individual programmes.[18,20] Depending on their TUG performance time, participants are separated into either a high or low risk of falling.[46] A participant could therefore do this test with two separate programmes that use different TUG cut-off times and be assessed as both high and low risk. Not having a universal cut-off time seemingly defeats the purpose of this test as the test should be applicable for any programme to use as an indicator of falls, with no prior knowledge of the participants.

4 Staged Balance Test

FSBT involves participants standing in four different stances for a maximum of 10 seconds. The score given is equivalent to the seconds achieved in each stance. The four stances include parallel, semi-tandem, tandem and one-legged.[47,48]

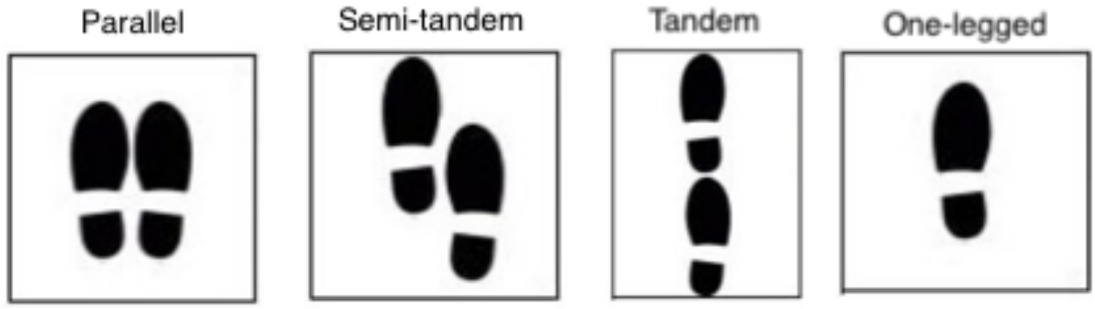


Figure 1. Four staged balance test feet positioning [47]

Many organisations recommend the FSBT is for assessing geriatric balance globally.[49] In the past this test has shown to have strong reliability,[50] as well as good intra- and inter-rater reliability.[51,52]

Internal validity has been established on the basis that participants' scores reflect their balance ability when compared with performance on other balance tests.[11,51,52] Although there is no research regarding the external validity of the FSBT in regard to falls risk, we could assume it might be low, due to research showing that many extrinsic risk factors for falling cannot be replicated in such tests.[9–11] Since the FSBT is conducted in a lab, it eradicates all external factors and it is therefore hard to evaluate whether it is applicable in the real-world.

30 Second Chair Stand

30 CST is a test which has been commonly used to measure leg strength and stability.[53] The aim of this test is to move from sitting to standing as many times as possible in 30 seconds.[54] The 30 CST has shown strong reliability through test-retest scenarios (ICC 0.84-0.92).[21,55]

Although the 30 CST may be a valid measure of leg strength it does not necessarily correlate with one's balance ability.[21,22] Macfarlane et al.[22] examined validity of the 30 CST by comparing scores with hip and knee extension strength. The 30 CST proved to be a moderately weak measure of lower leg strength. As well as having low validity, there is also contradicting research about its measure of leg strength.

The 30 CST also does not offer a universal scale to assess someone's fall risk, which therefore makes it hard for programmes to identify whether someone is high or low risk.

Balance Error Scoring System

The Balance Error Scoring System (BESS) test involves participants standing in three different test positions for 30 seconds. These positions include parallel, tandem and one-legged. This test is usually performed on both hard flooring and a foam pad.[56]

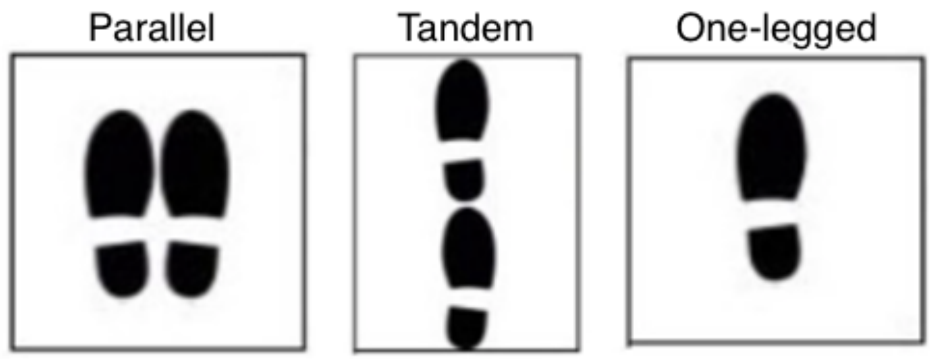


Figure 2. Balance error scoring system feet positioning [47]

The BESS test is scored by the number of errors accumulated in the 30 second timeframe. Participants are expected to have their eyes closed and hands on their iliac crests for the duration of the test. Potential errors include moving their hands off their iliac crests, opening their eyes, stepping, stumbling or falling, abduction of the hip beyond 30 degrees, lifting of the forefoot or heel, or remaining out of the test position for longer than five seconds.[56] If a participant commits multiple errors simultaneously it only counts as one error. The maximum total number of errors a participant can score per stance is 10.[56]

The BESS test has been shown to provide reliable measures of balance.[23–25,56] Although this test is reliable there is the question of its validity. Although a test can measure balance, this does not necessarily mean that it is an indication of someone's falls risk. As the test is conducted fully attention invested with no distractions, such conditions would be highly unlikely in the real world. Falls occur due to both intrinsic and extrinsic factors. Conducting BESS in a lab-style setting isolates only the intrinsic influences on falling.[7,8,10,11] These conditions are not comparable to uncontrolled environments, which raises questions concerning the external validity of such tests.[26]

Although these different balance tests may be reliable, a key consideration for these tests is their validity. Different tests can measure one or more different components of balance, but this does not necessarily mean these tests translate into someone's potential to fall. As discussed previously there are many intrinsic and extrinsic factors that have an effect on older adults falling. All of these tests tend to isolate the intrinsic factors and therefore do not incorporate any extrinsic potential. When performing in a lab-based setting with optimal environmental conditions, a participant can compensate for their usual ability by bringing more attentional focus to the task.[26] This means that participants may be balancing better in lab-based tests than in the real-world due to being able to focus solely on the balance task in the lab.

In life, our sensory, motor and cognitive elements constantly interact for all tasks we perform. It therefore seems logical that this same interaction should be replicated with test methods that are trying to replicate real-world scenarios.[26] The idea of adding a cognitive task to a balance test should create a more realistic and applicable indication as to whether someone is at risk of falling. This concept is known as divested attention training.

Divested Attention Training

Divested attention training is the idea of focusing on multiple tasks at one time.[27] The idea of divested attention training is that the performance of one task interferes with another task being performed simultaneously.[28] By performing these simultaneously, the participant's attentional demand is forced to spread or move between them.[29] This is particularly difficult when motor performance is not automated.[29]

There are two key theories about divesting attention, these are the capacity theory and the resource theory. The capacity theory believes everyone has a limited amount of attentional resources that can be used at one time, therefore by a primary task demanding the most resources it leaves little capacity for other concurrent tasks.[57,58] By performing more than one task the resources can be split resulting in both tasks not gaining maximum resources. Although this could be considered a negative, everyone's attentional capacity is fluid and can be easily transferred.[57,58] Although similar to capacity theory, resource theory highlights the allocation of attentional resources. It believes that one task will dominate the attentional resources as there has to be a primary task. This means although multitasking can happen, one task will be demanding the main proportion of one's attention. We would therefore see someone demonstrate an imbalance of ability across multiple simultaneous tasks.[57,58] Although similar, capacity theory views attention being more fluid and transferable compared to resource theory.

Importance of divested attention training

When testing a single-task skill it is arguably important to incorporate divested attention training, as many real-life scenarios will include participants having to split their attention across a range of tasks. By isolating a task it creates ample opportunity for a participant to solely focus on that task and use more attention than they normally would in everyday life.[59,60] Divested attention training can be used across a wide range of movement patterns including professional sport or everyday movement such as standing and walking. Divested attention training has shown to be successful for scenarios involving limb amputation, injury rehabilitation and improving movement among older adults.[28–31]

Research conducted among the older adult population studied the effect of divested attention training on movements such as the gait cycle and posture. No research was located examining the effect of divested attention training and balance ability. As individuals age their functioning begins to deteriorate; this includes their sensory, motor and cognitive processes, all of which play a part in everyday movement. It is therefore understandable that as these processes age and become impaired, that tasks will require more attentional demand.[26,27] Movements such as walking and maintaining posture would require more attention, therefore making multitasking more difficult, but still remaining a vital and common part of everyday life. Creating exercises and tests that mimic divested-attention training could be beneficial for older adults as it would help them apply this training to their real life. Through practicing under these circumstances, training may help decrease the attentional demand needed for everyday tasks.

Evidence of success of divested attention training

Silsupadol et al.[61,62] found that divested attention conditions provide more accurate information compared to a single-task conditions. When analysing gait cycles, five out of 32 studies used a memory or mental arithmetic task simultaneously with assessing gait. These five studies produced the most accurate and reliable results; and the other 27 studies were suggested as lacking relevance as environmental conditions were not relatable.[26] Other studies have found age and physical condition played a factor in whether single or dual task conditions affected the participants' performances. Studies that looked at postural balance found that a cognitive task decreased the quality of posture in older adults but had little effect on young adults.[28,63] One study in particular conducted research on a team of baseball players. This involved the players performing the BESS test in a lab-based setting and on the side-line during practice. The side-line scores were significantly higher (more errors were scored) due to the external stimuli of watching practice. This showed that a split in the attentional demand occurred between balancing and watching practice compared to the lab-based test where there were no distractions.[64] People with recent lower limb amputations also showed a larger disturbance in posture once distracted.[61,62,65] These studies show that the more subconscious the task is to perform, the less effect a subsequent task could have on performance. For example, holding posture could be more conscious in older adults due to their loss of muscle mass and strength, as well as the slowing of cognitive processes that occurs with older age.[11,66]

Divested attention training and balance ability

Divested attention training among older adults may be important for improving balance as falls occur more frequently during activities where attention has been divided.[28] When looking at divested attention training among older adults either the physical or mental task performance will be compromised and result in a decrease in performance of either one or both of the tasks.[32,33]

Agmon et al.[32] found that postural sway occurred once the participants began a cognitive task indicating that attention is needed for posture among older adults. Posture is assumed to require little to no attention as it is a well-learned skill, but this may not be the case for older adults due to physical deterioration.[33] One study showed that posture decreased for older adults when performing a cognitive skill but not for young adults, meaning that with age our regulation of posture becomes more challenging, meaning more attention is needed for control.[63] Divested attention training can therefore help assess whether someone's postural control has deteriorated and higher attentional demand is required. Single-task training is unlikely to measure this due to there being only the one task to focus on.[63,67]

Scherder et al.[30] found that some participants' balance will be consistent but their cognitive test scores will decrease. This showed that primary attention is given to balancing, which resulted in less attention being used for the cognitive task. A popular test to measure older adult's balance ability is the walking-while-talking test. An individual is filmed walking twice, once while having a conversation and the other without. Differences in the participants' gait cycles and pace will determine their risk of falling.[30,68]

Reliable testing

When conducting research, it is important that the data being gathered is reliable and valid. Some key areas for concern with divested attention training is the learning effect, the reliability of both tests that are being combined and whether it has strong inter-rater reliability and can be accurately scored by assessors.

Learning effect

The learning effect, also known as practice effect, is when an individual's performance improves due to repeated exposure to testing material resulting in them gaining familiarity and adopting strategies to improve.[69,70] A learning effect will be particularly noticeable in the early exposures to a test and usually does not occur later.[69] If the learning effect isn't eliminated through prior familiarisation, then data may not be reliable as it will have an increased variability of results.[69,71,72] For some tests the learning effect can last up to four sessions and therefore four prior sessions may need to be conducted to remove this variable.[70–72] It has also been found that tests with larger cognitive demands have the biggest learning effects.[70]

A BESS test-retest trial found that learning was obvious between the first and second sessions, however scores stabilised for subsequent sessions.[25,73] Other research found that there was no difference between the 30 day pre- and post-test scores of the BESS, but the average participant scored lower (less errors) on the tests five and seven days after pre-testing. This shows that the learning effect occurred within the first week but did not make a difference to post-testing 30 days later.[74] BESS performance therefore improves as a result of learning effects as participants develop strategies to better their performance.[75]

Reliability testing

The BESS test has been shown to be a reliable measure of balance and postural control, especially among older adults.[25] A number of studies found that BESS test-retest demonstrated good reliability, with results having an ICC of 0.74.[75,76] A study conducted on a group of American collegiate football players found that the control group only had a 5% difference in their test-retest scores, indicating strong reliability.[77]

The reliability of the COWA test was assessed by the total number of words spoken in 20 seconds. The reliability was 83% which is acceptable considering the test has a brief duration.[78,79] This study also assessed test-retest to assess reliability with two sessions. This had a moderate reliability of 74%.[78,79]

Inter-rater reliability

Numerous studies have evaluated inter-rater reliability for the BESS test. One study found that the BESS test had strong inter-rater reliability (ICC=0.81).[24] Two other studies found ICC reliability scores of 0.92 and 0.61 which showed that the inter-rater reliability of the BESS test is good to excellent.[75,76]

Summary

The available literature was reviewed to inform the methods of this study. The balance test used was a modified BESS, the BESS test was found to be the most consistent and reliable of the balance tests assessed. This test was combined with the COWA test, to create a dual-task test that divested the participants attention.

The reason for this was to replicate the real-world situation of having to balance while simultaneously performing other tasks or being distracted.

Chapter 3: Methods

Due to Covid-19 the revised aim of this study was to investigate the inter-rater reliability in scoring of the mBESS. The specific objective was to determine the inter-rater reliability between raters scoring the modified BESS test.

This chapter explains how inter-rater reliability of the mBESS was assessed. It will describe recruitment and participant eligibility, ethics and the study design and procedures. Lastly this section will show how the data was analysed.

Methods

Study design

To address the objectives of this study a test-retest protocol of the mBESS was used. Ethical approval for this study was obtained from the Otago Polytechnic Research Ethics Committee (21st February 2020) (Appendix 4). Kaitohutohu Research consultation was undertaken (8th November 2019) with the department's Kaitohutohu representative.

The first phase of the study involved testing participants over three testing sessions using the combined BESS and COWA test, which is referred to as the mBESS. Sessions were conducted over a two-week period, with each session being at least 3-4 days apart. In an initial familiarisation session, participants had the tests explained and were able to practice the tests and understand what the following sessions would entail. The three testing sessions were then conducted and recorded for scoring purposes. No intervention or training was provided within the two-week testing period. Protocols for the BESS test were based on the BESS manual,[56] and protocol for the COWA test was based on the Benton COWA Test: Reliability and Updated Norms.[78,79]

The second phase involved four independent raters individually scoring all participant tests using the criteria provided. The independent raters were purposively selected to provide a range of experiences and knowledge. They included individuals with a background in personal training, physiotherapy and occupational therapy, as well as an individual with no background in health and fitness. This was the first time all four raters had involvement with balance testing. Video recordings of testing sessions were provided to these raters.

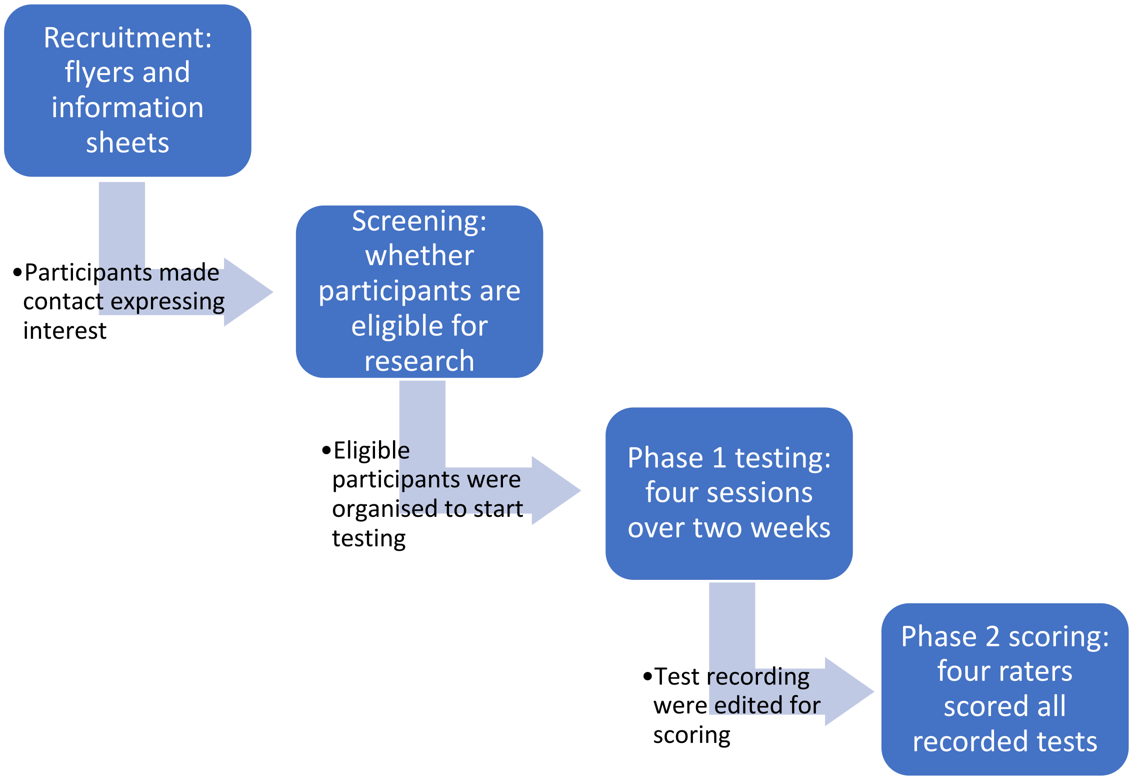


Figure 3. Recruiting and testing process

Participants

Participants were recruited from a local older adults exercise programme. Two forms of screening were employed with this study. When recruiting both the information sheet and flyer outlined the eligibility criteria. This gave participants the option to communicate with the researcher regarding any health concerns or queries. Each participant then completed a screening form that checked their eligibility. Basic information such as contact details, an emergency contact, lifestyle questions and medical history were captured. The lifestyle questions focused on how active the participant was and their falls history. The medical history focused on existing or previous health issues.

The eligibility criteria were intended to ensure the safety of the participants. As this is an inter-rater reliability study, participants were required to be over 18 years of age and speak fluent English. Participants were excluded if they had a current heart condition, had any known balance disorders, or had fallen in the last two years.

The four participants were all recruited from a local older adult exercise programme. They were all over 60 years old, three of which were males and one female.

Data collection procedures

At each testing session participants were asked to remove their shoes and stand in a designated area in the room. A spotter was positioned either side of the participant. Spotters were trained according to the British Columbia Institute of Technology[80] on safe patient handling, position and care guidelines. These guidelines include patient risk analysis, how to identify someone about to fall, and how to effectively help someone who is falling with correct posture, stance, hand placement, and technique.[80] A script of instructions was then read to the participant with verbal and visual instructions.

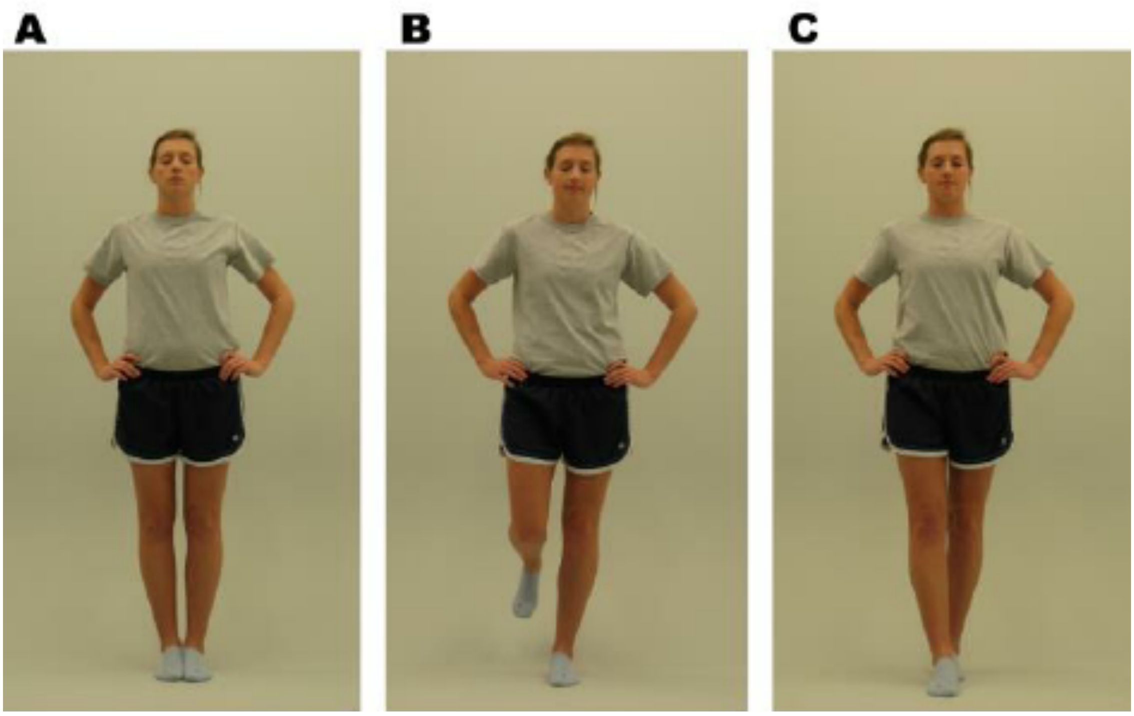


Figure 4. BESS test positions (parallel, one-legged & tandem) [56]

Participants were read and shown the modified BESS test instructions according to the University of North Carolina's[56] BESS manual. The modified COWA test was then explained following the COWA guidelines provided by the Benton COWA Test: reliability and updated norms and instructions [78,79] with minor modifications. These modifications were that instead of having one minute, the participant was provided the same amount of time as the mBESS test (20 seconds). Instead of only naming two words, participants were asked to list as many words as possible in the given timeframe. The letters provided were all English letters with exception of ‘q’ and ‘z’ and were randomly generated using a randomiser website (www.random.org).

At session one, participants were informed that it was a familiarisation session and that test results would not be recorded. For the following three sessions participants were reminded that their tests and test scores were recorded. Once the examiner communicated to the participant the assigned letter, the participant commenced balancing and reciting words beginning with that chosen letter. The timer started once the participant recited their first word. Participants were provided a 30 second rest between each test. During each session participants completed nine tests, three in each of the three positions. The practice session and three test sessions occurred over a two-week period, three to four days apart.

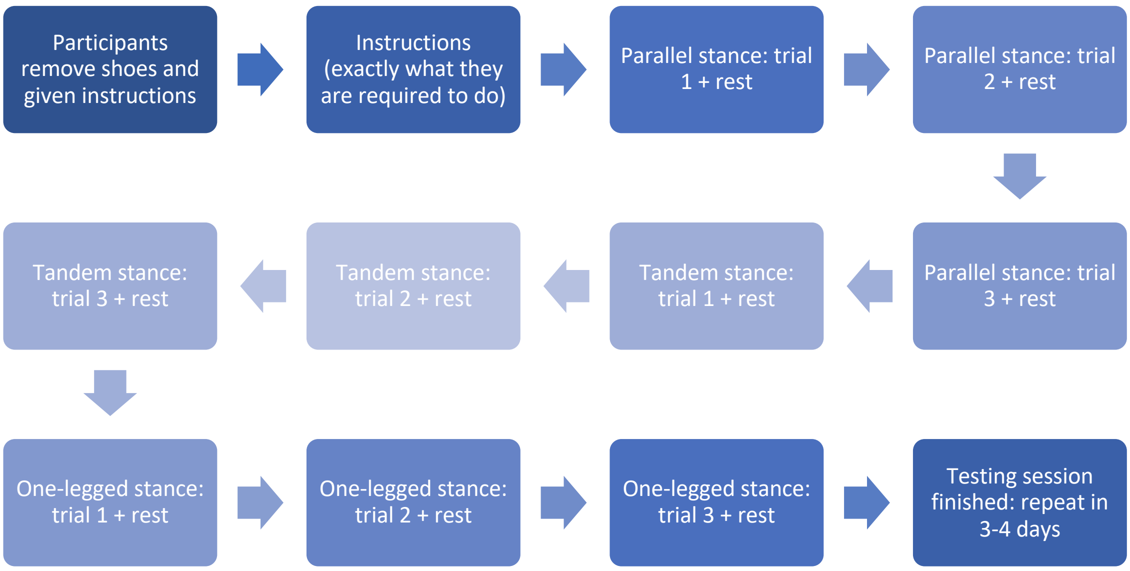


Figure 5. Participant testing process

Raters scoring

The independent raters were provided with a template scoring sheet that they were asked to complete for each test that they scored (Appendix 2). They were provided with an instruction sheet (Appendix 3), the BESS manual and four videos of the participants.

Each rater was asked to read through the instruction sheet and the BESS manual to understand what they were scoring. This explained that each participant had three recorded sessions, containing three trials in each of the three positions. This therefore meant each edited video contained 27 trials to score (3 trials in each of 3 sessions for 3 stances), a digital timer was visible on each video to help the raters with the 20 second test duration. The raters were instructed to score an error if:

- The participant's hands lifted from their hips
- The participant opened their eyes
- The participant stepped, stumbled or fell
- The participant abducted their support hip by more than 30 degrees
- The participant lifted their forefoot or heel
- The participant remained out of test position for five or more seconds

Clarifications were made that only one error was to be marked if two or more of the listed errors happened simultaneously and that one error was to be scored until the participant regained control. The raters were also told to ignore the scoring cap of ten errors. Each rater then individually scored all tests for the four participants and recorded their scores on the template provided.

Data analysis

All raw data was gathered through filming the tests. These tests were then edited into short films for the raters to watch. Each participant film included a total of 27 balance attempts from all three sessions. The attempts were edited together, a virtual timer was placed on the screen and the film was muted to make the rating process simple for the raters. To evaluate inter-rater reliability Pearson's correlation coefficient (PCC) was used. This compared the rater's scores against one another to assess inter-rater reliability. PCC was chosen as the research was trying to find the correlation between each two raters.

Chapter 4: Results

The initial intent of this research was to examine the test-retest reliability of the mBESS. However, due to Covid-19, testing was suspended with only four participants having completed all of their scheduled testing sessions. This resulted in insufficient statistical power for evaluating the test-retest reliability.

The modified research objective was to determine the inter-rater reliability making use of the limited data captured prior to lockdown. Four volunteers were recruited to independently rate the video recordings with these results statistically analysed. The four raters had varying careers in occupational therapy, physiotherapy, personal training and an irrelevant field to health and fitness.

The individual scores (tables 1 & 2) show that participants made more errors as the difficulty of the balance position increased. The initial scoring (the researcher) for the four participants indicated that the parallel stance (0-1 errors) and tandem stance (0-6 errors) had relatively few errors and low variability, while the one-legged stance demonstrated more errors and higher variability (5-12 errors). For most participants the parallel stance tests proved easy as zero errors were recorded. The variability of the scoring increased along with the difficulty of the stances (Table 2). The greatest scoring variability was noted for participant 3 during the one-legged stance, with scores varying among raters by 14.47%.

Table 1. Mean and range (of 3 trials) of errors scored by raters for each participant in each stance

	Participant one			Participant two			Participant three			Participant four		
	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged
Rater 1	0.2 (0-1)	2 (0-5)	5.7 (5-7)	0 (0)	0.4 (0-2)	6.8 (5-8)	0 (0)	0 (0)	9.7 (5-13)	0 (0)	0.6 (1-2)	6.8 (4-12)
Rater 2	0 (0)	1.9 (0-5)	5.8 (5-7)	0 (0)	0.2 (0-2)	7.1 (5-10)	0 (0)	0 (0)	9.4 (5-13)	0 (0)	0 (0)	7 (4-11)
Rater 3	0 (0)	2.1 (0-7)	5.9 (5-7)	0 (0)	0.1 (0-1)	7.1 (6-8)	0 (0)	0 (0)	8.4 (5-12)	0 (0)	0.1 (0-1)	7.4 (6-10)
Rater 4	0 (0)	2.1 (0-6)	5.8 (5-7)	0 (0)	0.1 (0-1)	6.8 (6-7)	0 (0)	0 (0)	9.3 (5-12)	0 (0)	0.3 (0-1)	7 (4-12)
Overall all	0.1 (0-1)	2.0 (0-7)	5.8 (5-7)	0 (0)	0.2 (0-2)	6.9 (5-10)	0 (0)	0 (0)	9.2 (5-13)	0 (0)	0.3 (0-2)	7.1 (4-12)

Scores for parallel stance were generally zero errors, with the exception of one rater scoring one error for one trial, so inter-rater reliability could not be evaluated statistically. The inter-rater reliability for tandem (Table 3) and single legged (Table 4) stances were strong, with reliability ranging from 0.77 – 0.97.[81,82]

Table 2. Inter-rater reliability of tandem stance

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.863	1		
Rater 3	0.868	0.967	1	
Rater 4	0.899	0.962	0.974	1

Table 3. Inter-rater reliability of one-legged stance

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.771	1		
Rater 3	0.778	0.861	1	
Rater 4	0.945	0.808	0.822	1

The mBESS test had excellent inter-rater reliability across all balance stances, averaging an overall inter-rater reliability of 0.972 (Table 6).

Table 4. Overall inter-rater reliability across all balance stances

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.959	1		
Rater 3	0.962	0.977	1	
Rater 4	0.986	0.971	0.980	1

The most challenging balance stance appeared to be the single legged stance. Radar plots (figures 4 to 7) demonstrate how each rater scored each trial for this stance.

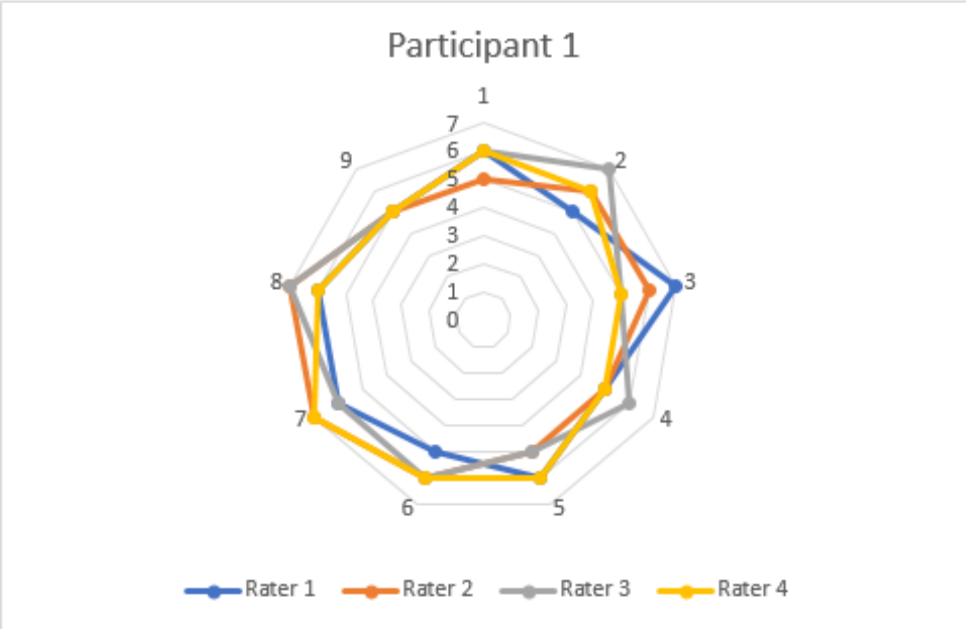


Figure 6. Errors scored one-legged stance, participant 1

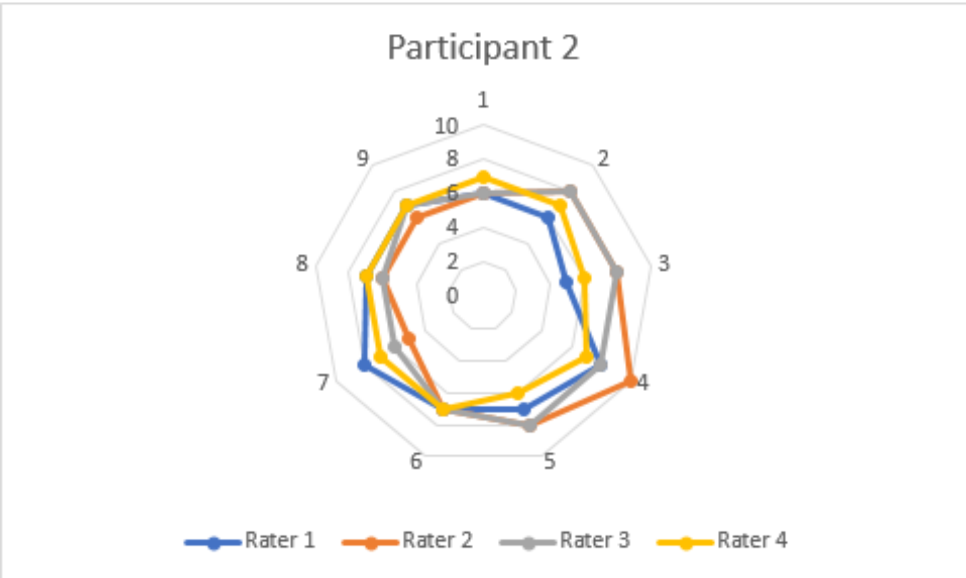


Figure 7. Errors scored one-legged stance, participant 2

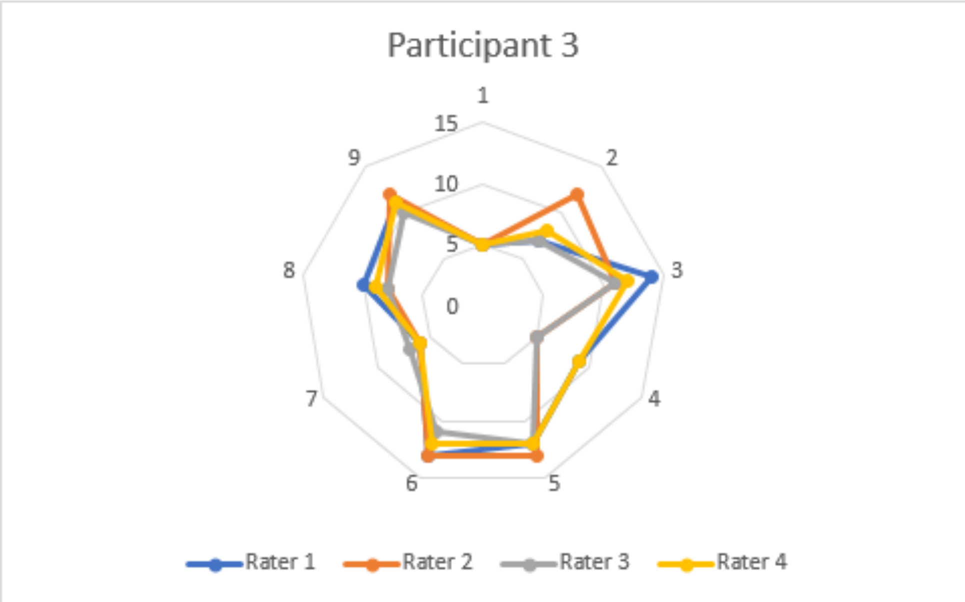


Figure 8. Errors scored one-legged stance, participant 3

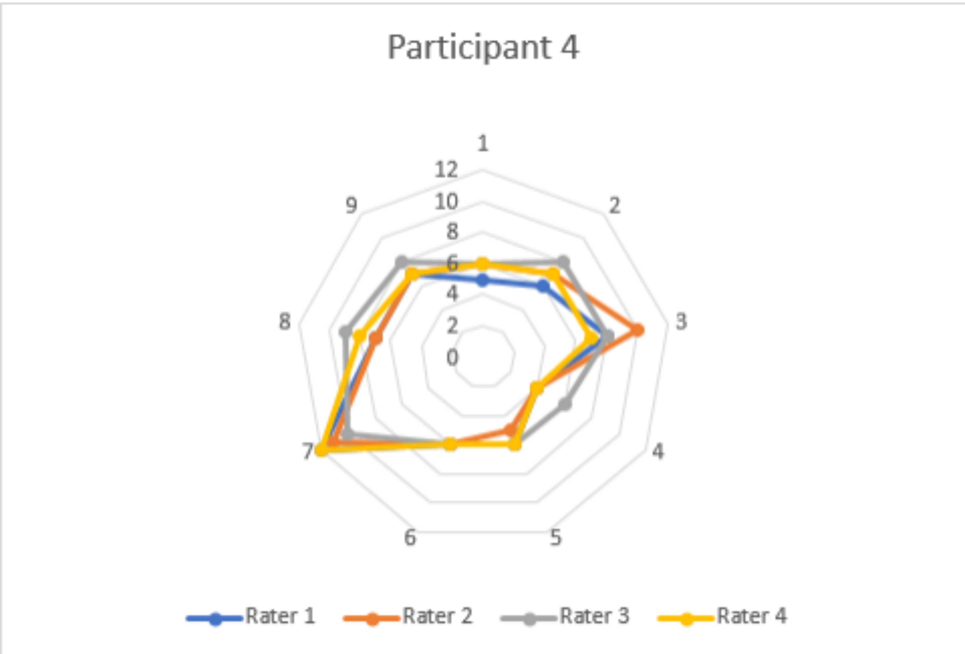


Figure 9. Errors scored one-legged stance, participant 4

Chapter 5: Discussion

This study explored a novel combination of a balance test (BESS) with a distracting cognitive task (COWA), with the expectation that the increased attentional demand needed for a secondary task would offer a more valid assessment of balance. The objective of this study was to examine inter-rater reliability of the new test (mBESS).

The mBESS had excellent inter-rater reliability. All stances had inter-rater reliability results greater than 0.75, which according to Cicchetti[81,82] is considered excellent. Although the BESS test protocol offers seemingly objective guidelines for scoring tests, it was found that as the number of errors increased the inter-rater reliability decreased due to more margin for variance in subjective ratings. Inter-rater reliability also decrease during the one-legged stance due to the differing balance strategies adopted by participants and more opportunities for multiple simultaneous errors. Both of these factors made it more challenging for the raters to score tests due to having to make subjective assessments of when a participant regained balance control.

According to Cicchetti[81,82] both the tandem and one-legged stance for the BESS are considered to have excellent inter-rater reliability. In this study employing a novel version of the BESS, similar inter-rater reliability was evident. With very few errors evident, the parallel stance could not be statistically evaluated for inter-rater reliability (34 out of 36 trials had zero scores). The tandem stance task had excellent inter-rater reliability, with correlation coefficients between 0.862 and 0.974. The one-legged stance had slightly lower inter-rater reliability, with scores ranging from 0.770 to 0.945. Due to the one-legged stance having an overall higher average of scores, with scores ranging from 51 to 87 it was expected there would be more variance compared to tandem stance due to the greater number of errors occurring. The high inter-rater reliability demonstrates that the mBESS test is capable of being reliably scored by any trained individual.

Previous research shows varying results regarding BESS and inter-rater reliability. Kleffelgård[24] found that the BESS had excellent inter-rater reliability, with an ICC of 0.86.[81,82] Another study by Finnoff et al.[83] found the BESS test to have an ICC reliability of 0.57, which is considered to be reliable.[81,82] From supporting research, the BESS seemed to have strong inter-rater reliability.[24] The current research supports this observation by demonstrating high inter-rater reliability.

Although the aims of this research were modified due to circumstances beyond the researcher's control, this did not change how the test was conducted with the participants. Explaining the methods and testing to participants was uncomplicated and participants appeared to understand the mBESS test requirements and procedures with few questions. All sessions were conducted seamlessly, and both the researcher and participants found the testing easy and straightforward to perform. Informal feedback from participants was that the test was novel and simple to follow. They suggested that the test was more difficult to perform than anticipated due to the cognitive distraction of the COWA test.

The idea of divested attention training is worth exploring with older adults. Their sensory, motor and cognitive processes deteriorate making everyday movements more difficult. More attentional demand is required for simple movements such as walking and posture, therefore multitasking becomes more challenging.[26,27] Both the capacity and resource divested attention theories suggest that everyone has a limited amount of attentional resources that can be used at any given time. Undertaking multiple tasks simultaneously will therefore create a split in attention leaving some tasks with less focus than others.[57,58] It is therefore understandable that the increase in attentional demand of a task could lead to losing balance and falling due to the shift in focus. Divested attention training could therefore arguably benefit older adults to help train their sensory, motor and cognitive processes, as this could help reduce their falls risk. The theory of divested attention training has already proved successful among those with limb amputations, injuries and limited movement.[28–31]

Strengths

There were two major strengths of this study, which were the strong inter-rater reliability and the innovative nature of the mBESS test. Across all three balance positions the inter-rater reliability was high[81,82] with the lowest inter-rater reliability being for the one-legged stance (0.77). The overall inter-rater reliability ranged from 0.959 – 0.985, which is extremely strong.

Another strength was the innovation of the test. With age comes cognitive and physical delays and impairments, which means balance tasks need more attention. It therefore seems a logical and relatively simple idea to test older adult's balance in a more ecologically valid way by combining balance with a distracting cognitive task. A combined test would offer more relevance for real-world situations due to the focus needed for balance and therefore prevent falling among older adults.[28] The distraction section of the mBESS was the COWA test, this task shifted the attentional focus from balancing to a cognitive task.

Limitations

The biggest limitation for this study was New Zealand's nationwide lockdown in response to Covid-19 that occurred two weeks into testing. Due to the lockdown guidelines and participants being over 60 years of age, it was understood that additional data could not be collected for at least eight weeks. With no indication when the lockdown would end, this interrupted testing and no further data could be collected on-site. This also halted participant recruitment as there was no indication of when lockdown would be lifted and whether this was achievable in this study's timeframe. For practical and ethical reasons, testing could not be conducted via online platforms, so data collection ceased completely.

The decision was therefore made to stop recruiting and use the current data collected. Full data on four participants and partial data on six participants were collected before lockdown occurred. The research focus on the reliability of the mBESS had to be abandoned and instead the research changed focus to the inter-rater reliability of scoring the test. This now leaves a gap for future research which could include a reliability study or using this test for a longitudinal study.

An observation made during mBESS testing was the challenge of spatial awareness of the participants. Since their eyes were closed it was hard for them to know where they were in relation to the room, as well as positioning to place their feet correctly in tandem stance. This meant that if participants were to change their body orientation during their trial it could make scoring more difficult for the rater as they were now facing a different direction. Due to the fixed video camera positioning the view could be interrupted and therefore present an additional challenge to scoring. Some participants who lost balance in the tandem stance often struggled to return to the same tandem position; for the same reasons, having their eyes closed and their subsequent reduced spatial awareness. Although they may have believed that they were in the same position, their heels were often not positioned at the front of their opposite foot's toe. This limitation was hard to resolve without compromising the study, as eyes were required to be closed for the duration of each BESS stance.

The techniques adopted by each participant could also affect scoring. For example, during the one-legged stance one participant might place their foot on the ground and take their time to regain balance before lifting their foot again, whereas another might tap and raise their foot multiple times attempting to regain balance. Although both techniques may have taken the same amount of time, and had the same balance outcome, they could be scored differently. For the purposes of this study the decision was made that an error counted as a continuous error until the participant regained control. Although this helped solve the issue of varying techniques, it raised another issue of the rater's perception of when the participant effectively regained control of balance. As the inter-rater reliability was high it seemed that all raters had similar interpretations of error scoring.

The urgency of participants to regain balance was also noticeably variable. When working with older adults the speed that they are able to move can vary between individuals. During testing, participants perceived as fitter appeared to regain control promptly and therefore had more time in the balance position. Whereas slower responders tended to have less time in balance. The conundrum is that more time in balance increases the opportunity to lose balance and be scored an error. Slower participants could take much longer to regain their balance position and therefore have less time to create errors.

The purpose of developing these balance tests was to be able to assess individual falls risk. Although the mBESS test demonstrated high inter-rater reliability, the test-retest reliability has not been investigated and the association of scores with falls risk remains unknown. No scale was found in the literature to indicate a threshold for falls risk. Due to varying individual techniques, scores could vary between individuals and not show an accurate representation of their balance ability. A way to manage this could be by instructing participants on the techniques that could be used if they lose balance.

Chapter 6: Conclusions

This research found that the mBESS test has strong inter-rater reliability, with an average inter-rater reliability of 0.972 across all stances. Due to the enforced change in research methods many questions remain concerning the test's reliability, difficulty and suitability for falls prevention among older adults. The mBESS proved quick and easy to administer, and both the researcher and participants found the test instructions simple to follow. Due to the relatively few errors and low variability, it was hard to judge whether this test was reliable for measuring an individual's balance ability. Ideally further research will be conducted regarding this test, exploring its reliability and applicability in the real world. Further research needs to be conducted as the idea of divested attention training could be an important step into preventing falls among older adults. Through this research it is understood that the world has an ageing population, if we do not act on lowering the amount and harm of falls this could be a huge cost to society that could be potentially minimised by effective exercise interventions.

Implications and future research directions

Current balance tests could be critiqued for having poor external validity due to being lab-based, and there is an identified need for more functionally relevant tests of balance. It is understood that movement uses more attentional demand as individuals age and training and testing should reflect this. Instead of training that permits the participant's sole focus to be on their balance, there is arguably value in adding a multitasking element.[26,27] One way to translate the real-world scenario of multitasking into training and testing is through cognitive distractions. Theoretically this would help decrease the attentional demand needed for movements and therefore reduce an individual's falls risk. As such the mBESS test could be ideal for testing the falls risk of older adults, due to the dual-task element.

A comparison of the mBESS test and the BESS would determine how much attention is drawn away from balancing when performing the COWA test. The next logical step would be a pilot study of the mBESS before using it confidentially in clinical practice. Unfortunately, due to Covid-19 the reliability design proposed for this study could not be completed, but this would be the most suitable next step. After establishing the tests reliability, further research would be ideal to compare the effects of divested attention balance training and the current single-task balance training.

References

1. World Health Organization. Falls. WHO. https://www.who.int/news-room/fact-sheets/detail/falls. Published 2018.
2. Khawaja M, Thomas N. Population ageing in New Zealand. Statistics NZ.
3. Manlapaz DG. *Exergaming to Improve Balance and Decrease the Risk of Falling in Individuals with Knee Osteoarthritis*. Dunedin; 2019.
4. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World J*. 2001;1:14. doi:10.1100/tsw.2001.58
5. World Health Organization. *World Report on Ageing and Health*.; 2015.
6. Lawler D. Falls data. ACC. Retrieved 4 August, 2020.
7. Al Aama T. Falls in the elderly: Spectrum and prevention. *Can Fam Physician*. 2011;57(7):771-776.
8. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off." *JAMA*. 2010;303(3):258-266. doi:10.1001/jama.2009.2024
9. Cameron ID, Gillespie LD, Robertson MC, et al. Interventions for preventing falls in older people in care facilities and hospitals. *Cochrane Database Syst Rev*. 2018;9. doi:10.1002/14651858.CD005465.pub3
10. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012. doi:10.1002/14651858.CD007146.pub3
11. Karlsson MK, Magnusson H, von Schewelov T, Rosengren BE. Prevention of falls in the elderly: a review. *Osteoporos Int*. 2013;24(3):747-762. doi:10.1007/s00198-012-2256-7
12. Kannus P, Sievänen H, Palvanen M, Järvinen T, Parkkari J. Prevention of falls and consequent injuries in elderly people. *Lancet*. 2005;366(9500):1885-1893. doi:10.1016/S0140-6736(05)67604-0
13. Ministry of Health. Needs Assessmet and Support Services for Older People: what you need to know. 2011:5.
14. Live Stronger For Longer. Find a Strength and Balance Class Near You. ACC.
15. Borowicz A, Zasadzka E, Gaczkowska A, Gawłowska O, Pawlaczyk M. Assessing gait and balance impairment in elderly residents of nursing homes. *J Phys Ther Sci*. 2016;28(9):2486-2490. doi:10.1589/jpts.28.2486
16. Health Quality & Safety Commission. An integrated approach to falls in older people: what is your part? 2017:5.
17. NICE. Assessment and prevention of falls in older people. *Natl Inst Heal Care Excell*. 2013;14(6):658. doi:10.7748/nop.26.6.18.e586
18. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. *BMC Geriatr*. 2014;14(1):14. doi:10.1186/1471-2318-14-14
19. Drootin M. Summary of the updated american geriatrics society/british geriatrics society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc*. 2011;59(1):148-157. doi:10.1111/j.1532-5415.2010.03234.x
20. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: A systematic review. *J Nutr Heal Aging*. 2011;15(10):933-938. doi:10.1007/s12603-011-0062-0
21. Unver B, Kalkan S, Yuksel E, Kahraman T, Karatosun V. Reliability of the 50-foot walk test and 30-sec Chair Stand Test in total knee arthroplasty. *Acta Ortop Bras*. 2015;23(4):184-187. doi:10.1590/1413-78522015230401018

22. Macfarlane DJ, Chou KL, Cheng YH, Chi I. Validity and normative data for thirty-second chair stand test in elderly community-dwelling Hong Kong Chinese. *Am J Hum Biol*. 2006;18(3):418-421. doi:10.1002/ajhb.20503
23. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. *Sports Health*. 2011;3(3):287-295. doi:10.1177/1941738111403122
24. Kleffelgård I, Langhammer B, Sandhaug M, Lundgaard Soberg H. Reliability and validity of the balance error scoring system – BESS. *Physiotherapy*. 2015;101(May):e764-e765. doi:10.1016/j.physio.2015.03.3636
25. Hunt T, Ferrara MS, Bornstein RA, Baumgartner TA. The Reliability of the Modified Blance Error Scoring System. *Clin J Sport Med*. 2009;19(6):471-475.
26. Mulder T, Zijlstra W, Geurts A. Assessment of motor recovery and decline. *Gait Posture*. 2002;16(2):198-210. doi:10.1016/S0966-6362(01)00157-6
27. Swanenburg J, de Bruin ED, Stauffacher M, Mulder T, Uebelhart D. Effects of exercise and nutrition on postural balance and risk of falling in elderly people with decreased bone mineral density: randomized controlled trial pilot study. *Clin Rehabil*. 2007;21(6):523-534. doi:10.1177/0269215507075206
28. Swanenburg Favero, Uebelhart & Mulder de B. The reliability of postural balance measures in single and dual tasking in elderly falls and non-fallers. *BMC Musculoskelet Disord*. 2008;9(1):162. doi:10.1186/1471-2474-9-162
29. de Visser E, Veth RP, Schreuder HW, Duysens J, Mulder T. Reorganization of gait after limb-saving surgery of the lower limb. *Am J Phys Med Rehabil*. 2003;82(11):825-831. doi:10.1097/01.PHM.0000091981.41025.FC
30. Scherder E, Eggermont L, Swaab D, et al. Gait in ageing and associated dementias; its relationship with cognition. *Neurosci Biobehav Rev*. 2007;31(4):485-497. doi:https://doi.org/10.1016/j.neubiorev.2006.11.007
31. Dault MC, Geurts ACH, Mulder TW, Duysens J. Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations. *Gait Posture*. 2001;14(3):248-255. doi:10.1016/S0966-6362(01)00130-8
32. Agmon M, Belza B, Nguyen HQ, Logsdon RG, Kelly VE. A systematic review of interventions conducted in clinical or community settings to improve dual-task postural control in older adults. *Clin Interv Aging*. 2014;9:477. doi:10.2147/CIA.S54978
33. Dault MC, Geurts ACH, Mulder TW, Duysens J. Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations. *Gait Posture*. 2001;14(3):248-255. doi:https://doi.org/10.1016/S0966-6362(01)00130-8
34. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*.; 2010.
35. McPherson L. National Population Projections: 2014. Statistics NZ.
36. Henwood TR, Taaffe DR. Short-term resistance training and the older adult: The effect of varied programmes for the enhancement of muscle strength and functional performance. *Clin Physiol Funct Imaging*. 2006;26(5):305-313. doi:10.1111/j.1475-097X.2006.00695.x
37. Mittaz Hager AG, Mathieu N, Lenoble-Hoskovec C, Swanenburg J, De Bie R, Hilfiker R. Effects of three home-based exercise programmes regarding falls, quality of life and exercise-adherence in older adults at risk of falling: protocol for a randomized controlled trial. *BMC Geriatr*. 2019;19(1):13. doi:10.1186/s12877-018-1021-y

38. Killingback C, Tsofliou F, Clark C. Older people's adherence to community-based group exercise programmes: A multiple-case study. *BMC Public Health*. 2017;1:115. doi:10.1186/s12889-017-4049-6
39. Lamb SE, Jorstad-Stein EC, Hauer K, Becker C. Development of a common outcome data set for fall injury prevention trials: The Prevention of Falls Network Europe consensus. *J Am Geriatr Soc*. 2005;53(9):1618-1622. doi:10.1111/j.1532-5415.2005.53455.x
40. Jagnoor J, Suraweera W, Keay L, et al. Childhood and adult mortality from unintentional falls in India. *Bull World Heal Organ*. 2011;89:733-740. doi:10.2471/blt.11.086306
41. Feder G, Cryer C, Donovan S, Carter Y. Guidelines for the prevention of falls in people over 65. *BMJ*. 2000;321(7267):1007-1011. doi:10.1136/bmj.321.7267.1007
42. Rimland JM, Abraha I, Dell'Aquila G, et al. Interventions for preventing falls in older people in care facilities and hospitals (Review) Interventions for preventing falls in older people in care facilities and hospitals. *Jama*. 2017;12(8). doi:10.1001/jama.2017.15006
43. Karinkanta S, Heinonen A, Sievänen H, et al. A multi-component exercise regimen to prevent functional decline and bone fragility in home-dwelling elderly women: Randomized, controlled trial. *Osteoporos Int*. 2007;18(4):453-462. doi:10.1007/s00198-006-0256-1
44. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012;9. doi:10.1002/14651858.CD007146.pub2
45. Pollock AS, Durward BR, Rowe PJ, Paul JP. What is balance? *Clin Rehabil*. 2000;14(4):402-406. doi:10.1191/0269215500cr342oa
46. Health Quality & Safety Commission. The Timed Up and Go (TUG) Test. Health Quality & Safety Commission. https://www.hqsc.govt.nz/assets/Falls/PR/004-falls-toolkit-TUG-test.pdf. Published 2017.
47. Health Quality & Safety Commission. The Four Stage Balance Test. Health Quality & Safety Commission. https://www.hqsc.govt.nz/assets/Falls/PR/006-falls-toolkit-four-stage-balance-test.pdf. Published 2017.
48. Centers for Disease Control and Prevention. The 4-Stage Balance Test. CDC.
49. Centers for Disease Control and Prevention. Materials for Healthcare Providers. Centers for Disease Control and Prevention. https://www.cdc.gov/steadi/materials.html. Published 2019. Accessed June 9, 2020.
50. Franchignoni F, Tesio L, Martino MT, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. *Aging Clin Exp Res*. 1998;10(1):26-31. doi:10.1007/bf03339630
51. Goecke J. *Effectiveness of the Stepping on Program in Fall Prevention Measured by the Four Stage Balance Test (FSBT).*; 2017.
52. Nelson J. *The Effectiveness of the "Stepping On" Program for Reducing the Incidence of Falls in the Elderly: Measured by Four Stage Balance Test.*; 2016.
53. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthr Cartil*. 2013;21(8):1042-1052. doi:10.1016/j.joca.2013.05.002
54. Health Quality & Safety Commission. The 30-Second Chair Stand Test. HQSC. https://www.hqsc.govt.nz/assets/Falls/PR/005-falls-toolkit-chair-stand-test.pdf.

Published 2017.
55. Hopkins WG. Measures of reliability in sports medicine and science. *Sport Med*. 2000;30(1):1-15. doi:10.2165/00007256-200030010-00001
56. University of North Carolina. Balance Error Scoring System (BESS). 2010:1-5.
57. Wickens C. *Processing Resources in Attention: Varieties of Attention*.; 1991.
58. Young MS, Stanton NA. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Hum Factors*. 2002;44(3):2002. doi:10.1518/0018720024497709
59. Hunt C, Paez A, Folmar E. The Impact of Attentional Focus on the Treatment of Musculoskeletal and. *Int J Sports Phys Ther*. 2017;12(6):901-907. doi:10.16603/ijspt20170901
60. Swanenburg J, de Bruin ED, Uebelhart D, Mulder T. Falls prediction in elderly people: a 1-year prospective study. *Gait Posture*. 2010;31(3):317-321. doi:10.1016/j.gaitpost.2009.11.013
61. Silsupadol P, Shumway-Cook A, Lugade V, et al. Effects of Single-Task Versus Dual-Task Training on Balance Performance in Older Adults: A Double-Blind, Randomized Controlled Trial. *Arch Phys Med Rehabil*. 2009;90(3):381-387. doi:10.1016/j.apmr.2008.09.559
62. Silsupadol P, Siu KC, Shumway-Cook A, Woollacott MH. Training of balance under single- and dual-task conditions in older adults with balance impairment. *Phys Ther*. 2006;86(2):269-281. doi:10.1093/ptj/86.2.269
63. Swanenburg J, de Bruin ED, Uebelhart D, Mulder T. Compromising postural balance in the elderly. *Gerontology*. 2009;55(3):353-360. doi:10.1159/000212757
64. Onate JA, Beck BC, Van Lunen BL. On-field testing environment and balance error scoring system performance during preseason screening of healthy collegiate baseball players. *J Athl Train*. 2007;42(4):446.
65. Kim GY, Han MR, Lee HG. Effect of dual-task rehabilitative training on cognitive and motor function of stroke patients. *J Phys Ther Sci*. 2014;26(1):1-6. doi:10.1589/jpts.26.1
66. van Dijk H, Mulder T, Hermens HJ. Effects of age and content of augmented feedback on learning an isometric force-production task. *Exp Brain Res Aging Res*. 2007;33(3):341-353. doi:10.1080/03610730701319194
67. Woollacott M, Shumway-Cook A. Attention and the control of posture and gait: A review of an emerging area of research. *Gait Posture*. 2002;16(1):1-14. doi:10.1016/S0966-6362(01)00156-4
68. Verghese J, Buschke H, Viola L, et al. Validity of divided attention tasks in predicting falls in older individuals: A preliminary study. *J Am Geriatr Soc*. 2002;50(9):1572-1576. doi:10.1046/j.1532-5415.2002.50415.x
69. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci*. 2010;11:118. doi:10.1186/1471-2202-11-118
70. Beglinger LJ, Gaydos B, Tangphao-Daniels O, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. *Arch Clin Neuropsychol*. 2005;20(4):517-529. doi:10.1016/j.acn.2004.12.003
71. Wu G, Sanderson B, Bittner V. The 6-minute walk test: How important is the learning effect? *Am Hear J*. 2003;146(1):129-133. doi:10.1016/S0002-8703(03)00119-4
72. Hamilton DM, Haennel RG. Validity and reliability of the 6-minute walk test in a cardiac rehabilitation population. *J Cardiopulm Rehabil*. 2000;98(4):156-164. doi:10.1097/00008483-200005000-00003

73. McLeod TCV, Perrin DH, Guskiewicz KM, Shultz SJ, Diamond R, Gansneder BM. Serial administration of clinical concussion assessments and learning effects in healthy young athletes. *Clin J Sport Med*. 2004;14(5):287-295. doi:10.1097/00042752-200409000-00007
74. Valovich TC, Perrin DH, Gansneder BM. Repeat administration elicits a practice effect with the Balance Error Scoring System but not with the Standardized Assessment of Concussion in high school athletes. *J Athl Train*. 2003;38(1):51.
75. Broglio SP, Zhu W, Sopiarz K, Park Y. Generalizability theory analysis of balance error scoring system reliability in healthy young adults. *J Athl Train*. 2009;44(5):497-502. doi:10.4085/1062-6050-44.5.497
76. Clark RC, Saxion CE, Cameron KL, Gerber JP. Associations between three clinical assessment tools for postural stability. *N Am J Sports Phys Ther*. 2010;5(3):122.
77. Davis GA, Iverson GL, Guskiewicz KM, Ptito A, Johnston KM. Contributions of neuroimaging, balance testing, electrophysiology and blood markers to the assessment of sport-related concussion. *Br J Sport Med*. 2009;43(1):36-45. doi:10.1136/bjsm.2009.058123
78. Ruff RM, Light RH, Parker SB, Levin HS. Benton controlled Oral Word Association Test: Reliability and updated norms. *Arch Clin Neuropsychol*. 1996;11(4):329-338. doi:10.1016/0887-6177(95)00033-X
79. Ross TP, Calhoun E, Cox T, Wenner C, Kono W, Pleasant M. The reliability and validity of qualitative scores for the Controlled Oral Word Association Test. *Arch Clin Neuropsychol*. 2007;22(4):475-488. doi:10.1016/j.acn.2007.01.026
80. British Columbia Institute of Technology. Safe patient handling, position and trasnfers. 2018:144-148. https://opentextbc.ca/clinicalskills/chapter/3-7-fall-prevention/.
81. Cicchetti D V. Choice of Agreement Statistics: A Discussion of the Underlying Biostatistics, With Heuristic Examples From the Vineland™–3. *NCS Pearson*. 2017.
82. Cicchetti D V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol Assess*. 1994;6(4):284-290. doi:10.1037/1040-3590.6.4.284
83. Finnoff JT, Peterson VJ, Hollman JH, Smith J. Intrarater and Interrater Reliability of the Balance Error Scoring System (BESS). *PM R*. 2009;1(1):50-54. doi:10.1016/j.pmrj.2008.06.002

Appendices

Appendix 1: Manuscript for submission to Physical Medicine & Rehabilitation

Inter-rater reliability of the modified Balance Error Scoring System

Imbalance; inter-rater reliability?

Lily Purdon (BPhed)
Otago Polytechnic: Institute of Sport, Exercise & Health

No source of funding.
I'd like to acknowledge my academic mentors Phil Handcock and Richard Humphrey, as well as my participants who made this possible.

Abstract

Introduction: Falling is a common occurrence among older adults with over 1/3 of older adults falling per year. A range of balance tests are used to indicate an older adults risk of falling, many of which are attention invested and arguably do not apply to real-world settings. In real-world situations an individual's primary focus cannot always be fixed on maintaining balance. This study used divested attention by combining the balance error scoring system (BESS), an established balance test, with the controlled oral word association (COWA) test.

Objective: The objective of this research was to examine the inter-rater reliability of a modified BESS (mBESS) test.

Design: The mBESS test was administered to four older adults, each completing three recorded sessions involving three trials of each balance stance. These recordings were then submitted to four raters who scored the tests independently using the BESS scoring criteria provided. Inter-rater reliability was assessed using Pearson's correlation coefficients.

Participants: Four older adult participants complete the mBESS testing and four raters independently scored the mBESS tests.

Results: The mean inter-rater reliability was 0.97, with the lowest reliability observed when scoring the one-legged stance (r=0.83). As expected, the variability between raters increased with the difficulty of the balance stance task. The mBESS test therefore offers excellent inter-rater reliability.

Conclusion: The mBESS test has excellent inter-rater reliability and the test could be suitable for future longitudinal and cross-sectional research regarding falls prevention to help assess falls risk with more authenticity.

Keywords

Inter-rater reliability, balance, falls, prevention, elderly

Introduction

With the growing concern around elderly falls and the potential harm it can cause, this project explored the feasibility of a new test that could accurately evaluate an older adult's fall risk. This research was designed to examine the inter-rater reliability of a test developed by combining two existing tests, the Balance Error Scoring System (BESS) and the Controlled Oral Word Association (COWA) test. The combined test is referred to as the modified BESS test (mBESS). The aim of this study was therefore to investigate between rater reliability in scoring the mBESS.

Background

Old age is defined as anyone over 65 years of age.[3] With ageing comes psychological and physical deterioration, both of which can contribute to falling. [1–3] Falls are the second leading cause of accidental injury or death worldwide, with 193,974 falls claims from older New Zealanders in 2017 [4,5]with an estimated cost of over $267 million. [5]

Falls are often attributed to intrinsic and/or extrinsic factors. Intrinsic factors include any psychological, physical or physiological impairments, whereas extrinsic factors are environmental issues such as, poor lighting, uneven flooring, polypharmacy and poor footwear. [6–10] The strongest predictor of whether someone is likely to be at risk of falling is their history of falls. [7–10]

Falling can have serious consequences including fractures or tissue damage, functional limitations, decreased independence, and in the worst case, death. [10,11] Not only can this be a burden on the individual, but it also comes with a significant cost to society. [5]

Many exercise classes evaluate participants' balance ability and risk of falling with functional tests. Common tests used nationally and globally include the Timed Up & Go (TUG), the Four Staged Balance Test (FSBT), 30 Second Chair Stand (30 CST) and the Balance Error Scoring System (BESS).[12,13] These tests have all demonstrated good reliability, including intra- and inter-rater reliability. [14–17] Although TUG has commonly been used, there are concerns that it offers an inadequate prediction of an individual's risk of falling. [18–21]

Although all of these tests measure an aspect of balance, whether it is mobility, strength, muscular endurance or flexibility, they are all conducted in well controlled settings. All tests are performed without distractions and therefore arguably do not represent real-world balance demands.[22] This therefore raises the question as to whether these tests offer the external validity for assessing an individual's fall risk. Falls commonly result from multitasking or becoming distracted, and this ideally should be translated into testing to create a more externally valid test.[6–8,10]

Divested attention training, also known as dual-task training, requires focusing on multiple tasks at one time.[23] This form of training is used to create interference between two tasks being performed simultaneously[24], forcing the participant's attentional demand to be spread between tasks and therefore potentially altering performance of each task.[25]

Divested attention training been used successfully with participants rehabilitating, recovering from lower limb amputation, or having lost functional movement with ageing.[24–27] Research on older adults showed the benefits of divested attention training for gait and posture.[28,29] As falls often occur due to distractions and multitasking[24], divested attention training could be ideal for older adults, challenging performances of either or both of the tasks.[29,30]

Inter-rater reliability is an important aspect to explore during test development as good inter-rater reliability means that different testers reliably score tests performed at different times and/or different sites.[31] The inter-rater reliability of the BESS was shown to be excellent with an interclass correlation coefficient (ICC) of 0.81.[21]

Significance of the study

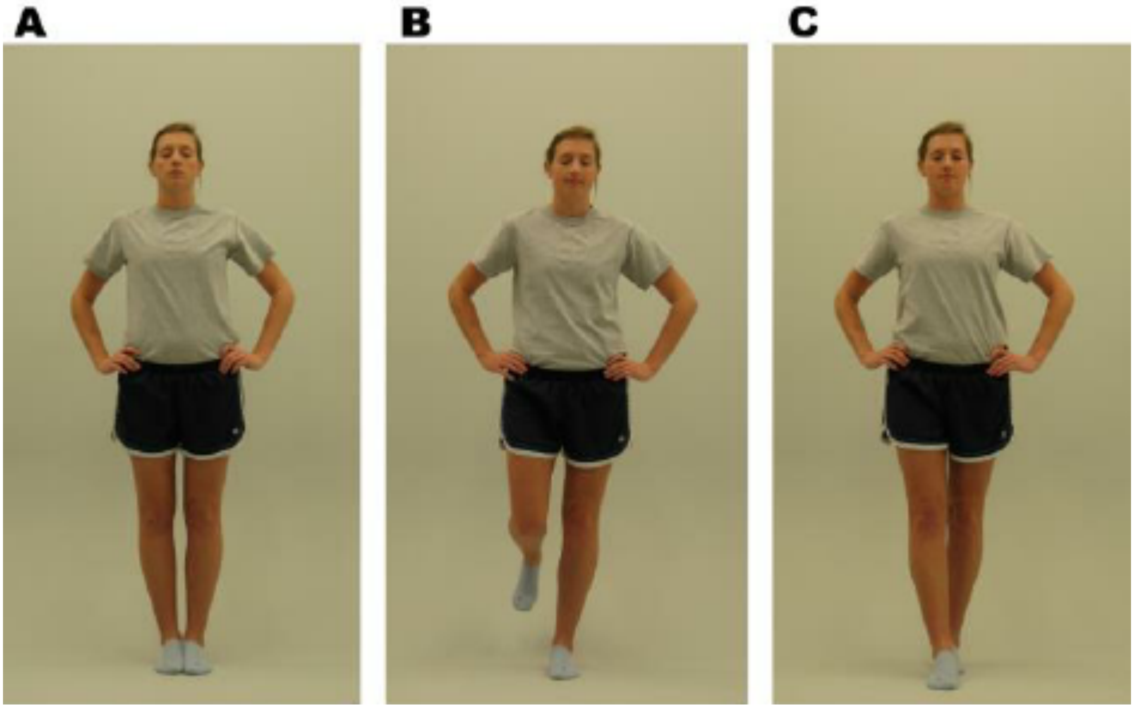
Falls are the second leading cause of injury or death worldwide and prevention strategies are important and require evaluation.[19] This research project combined a respected balance test with cognitive distractions and examined the inter-rater reliability of this test. Currently there is no known research that has explored divested attention training and older adults' balance ability. Due to its success with other populations[24–27], this research avenue could help suggest new strategies for older adults' falls prevention, help to reduce incidence and impact of falls.

Methods

Ethical approval for this study was obtained from the Otago Polytechnic Research Ethics Committee (21st February 2020). The study involved two groups; four volunteers who completed balance testing (participants), and four individuals who were invited to independently score the balance tests (raters).

Participants were familiarised with the testing procedures and attended three subsequent testing sessions where they perform three trials of each stance (parallel, tandem and one-legged). While performing the BESS test they simultaneously performed the COWA test, which involves listing as many words as possible beginning with the assigned letter.

Figure 10. Parallel, tandem and one-legged stance used in mBESS test [32]



The raters were purposively selected to provide a range of experiences and knowledge. (backgrounds in personal training, physiotherapy, occupational therapy, no health or fitness background). Video recordings of all (4 participants, 3 sessions x 3 stances x 3 trials) testing sessions were provided to the raters who were asked to score the mBESS trials using the BESS scoring protocol.[32]

They were provided with a template scoring sheet that they were asked to complete for each test. Raters were asked to read through the instruction sheet and the BESS manual to understand what they were scoring. Each edited video recording contained 27 trials to score (3 trials in each of 3 sessions for 3 stances), a digital timer was visible on each video to help the raters with the 20 second test duration. The raters were instructed to score an error if:

- Their hands lifted from their hips
- They opened their eyes
- They stepped, stumbled or fell
- They abducted their support hip by more than 30 degrees
- They lifted their forefoot or heel
- They remained out of test position for five or more seconds

Only one error was to be scored if two or more of the listed errors happened simultaneously and until the individual regained control. Raters were also told to ignore the BESS scoring cap of ten errors. Each participant then individually scored all tests and recorded their scores on the template provided. To evaluate inter-rater reliability Pearson's correlation coefficient (PCC) were calculated.

Results

The research objective was to determine the inter-rater reliability of the mBESS. Four raters independently rated video recordings of mBESS balance testing. Participants made more errors (tables 1 & 2) as the difficulty of the balance position increased. The variability of the scoring increased along with the difficulty of the stances (Table 2). The greatest scoring variability was noted for participant 3 during the one-legged stance, with scores varying among raters by 14.47%.

Table 1. Mean and range (of 3 trials) of errors scored

	Participant one			Participant two			Participant three			Participant four		
	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged	Parallel	Tandem	One-legged
Rater 1	0.2 (0-1)	2 (0-5)	5.7 (5-7)	0 (0)	0.4 (0-2)	6.8 (5-8)	0 (0)	0 (0)	9.7 (5-13)	0 (0)	0.6 (1-2)	6.8 (4-12)
Rater 2	0 (0)	1.9 (0-5)	5.8 (5-7)	0 (0)	0.2 (0-2)	7.1 (5-10)	0 (0)	0 (0)	9.4 (5-13)	0 (0)	0 (0)	7 (4-11)
Rater 3	0 (0)	2.1 (0-7)	5.9 (5-7)	0 (0)	0.1 (0-1)	7.1 (6-8)	0 (0)	0 (0)	8.4 (5-12)	0 (0)	0.1 (0-1)	7.4 (6-10)
Rater 4	0 (0)	2.1 (0-6)	5.8 (5-7)	0 (0)	0.1 (0-1)	6.8 (6-7)	0 (0)	0 (0)	9.3 (5-12)	0 (0)	0.3 (0-1)	7 (4-12)
All	0.1 (0-1)	2.0 (0-7)	5.8 (5-7)	0 (0)	0.2 (0-2)	6.9 (5-10)	0 (0)	0 (0)	9.2 (5-13)	0 (0)	0.3 (0-2)	7.1 (4-12)

As parallel stance produced zero errors, with the exception of one rater scoring one error for one trial, inter-rater reliability could not be evaluated statistically for this stance. The inter-rater reliability for tandem (Table 3) and single legged (Table 4) stances were strong, with reliability ranging from 0.77 – 0.97. [33,34]

Table 2. Inter-rater reliability of tandem stance

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.86264344	1		
Rater 3	0.86771283	0.96659233	1	
Rater 4	0.89895905	0.96164211	0.97449116	1

Table 3. Inter-rater reliability of one-legged stance

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.77092219	1		
Rater 3	0.77823393	0.86120346	1	
Rater 4	0.94481492	0.80771302	0.82238068	1

The mBESS test had excellent inter-rater reliability across all balance stances, averaging an overall inter-rater reliability of 0.972 (Table 6).

Table 4. Overall inter-rater reliability across all balance stances

	Rater 1	*Rater 2*	*Rater 3*	*Rater 4*
Rater 1	1			
Rater 2	0.95889857	1		
Rater 3	0.96183646	0.97708216	1	
Rater 4	0.98575818	0.97129512	0.97958288	1

The most challenging balance stance appeared to be the single legged stance. Radar plots (figures 4 to 7) demonstrate how each rater scored each trial for this stance.

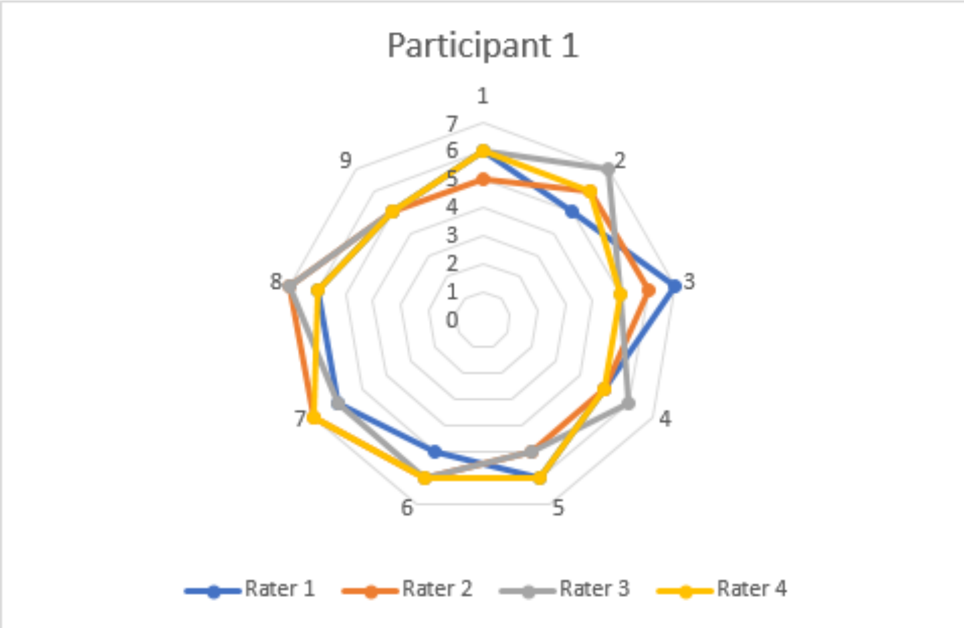


Figure 11. Errors scored one-legged stance, participant 1

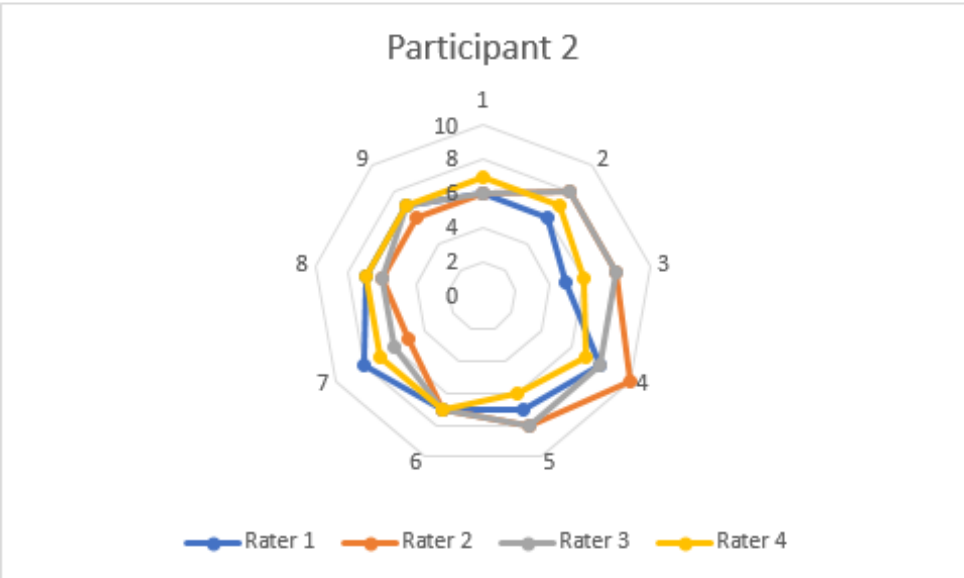


Figure 12. Errors scored one-legged stance, participant 2

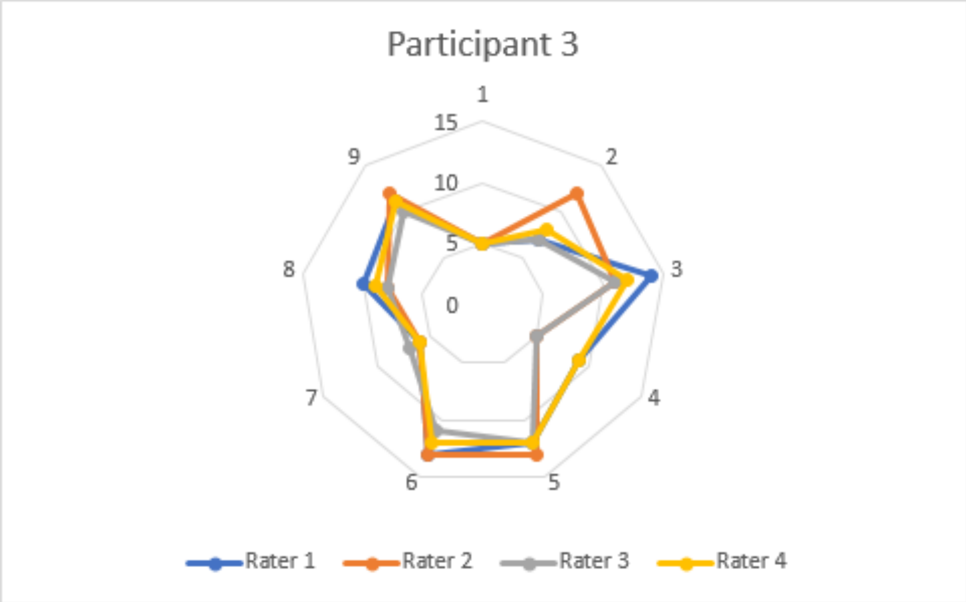


Figure 13. Errors scored one-legged stance, participant 3

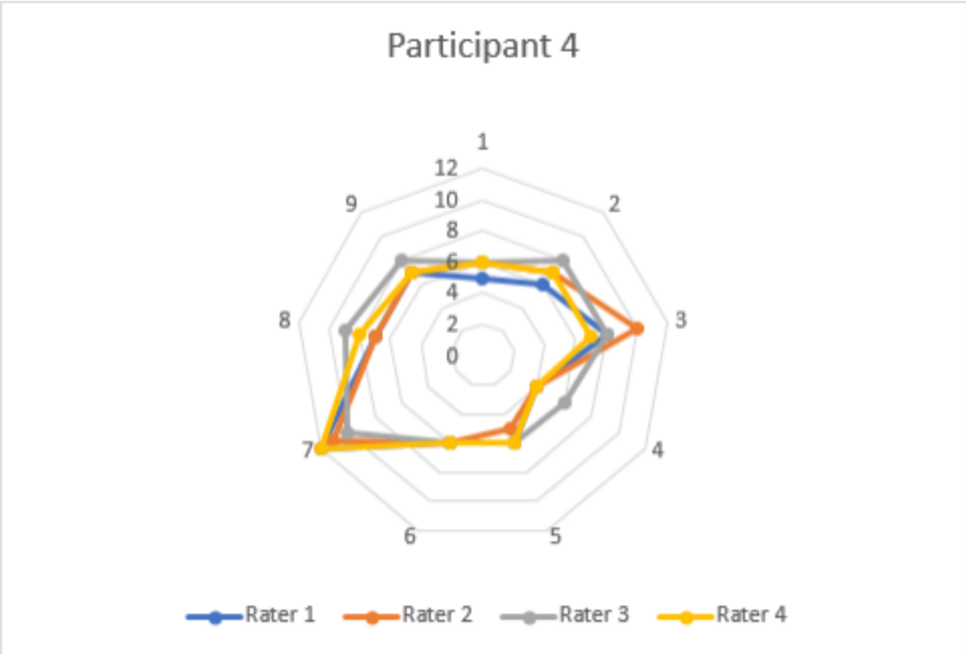


Figure 14. Errors scored one-legged stance, participant 4

Discussion

This study explored a novel combination of a balance test (BESS) with a distracting cognitive task (COWA), with the expectation that the increased attentional demand would offer a more ecologically valid assessment of balance. The objective of this study was to examine inter-rater reliability when scoring the new test (mBESS).

The mBESS had excellent inter-rater reliability. All stances had inter-rater reliability results greater than 0.75, which according to Cicchetti [33,34] is considered excellent. Although the BESS test protocol offers seemingly objective guidelines for scoring tests, it was expected that as the number of errors increased that inter-rater reliability would decrease due to more margin for variance in subjective ratings. Inter-rater reliability was also expected to decrease during the one-legged stance due to the differing balance strategies adopted by participants and more opportunities for multiple simultaneous errors. Both of these factors made it more challenging for the raters to score tests due the subjectivity of deciding when a participant had regained balance control.

According to Cicchetti [33,34] both the tandem and one-legged stance for the BESS test are considered to have excellent inter-rater reliability. For the novel version of the BESS employed in the present study, a similar inter-rater reliability was evident. Due to the one-legged stance having an overall higher average of scores, with scores ranging from 4-12 it was expected there would be more variance compared to tandem stance due to the greater number of errors occurring. However, the high inter-rater reliability from this study demonstrates that the mBESS test can be reliably scored by any trained individual.

Previous research has shown varying results regarding BESS and inter-rater reliability.[17] Kleffelgård[17] found that the BESS had excellent inter-rater reliability, with an ICC of 0.86.[33,34] Another study by Finnoff et al.[35] found the BESS to have an ICC reliability of 0.57, which is considered reliable.[33,34] The current research supports this observation.

Explaining the methods and testing to participants was uncomplicated and participants appeared to understand the mBESS test requirements and procedures with few questions. All sessions were conducted seamlessly, and both the researcher and participants found the testing easy and straightforward to perform. Informal feedback from participants was that the test was novel and simple to follow. They suggested that the test was more difficult to perform than anticipated due to the cognitive distraction of the COWA test.

The idea of divested attention training is worth exploring with older adults. Their sensory, motor and cognitive processes deteriorate making everyday movements more difficult. More attentional demand is required for simple movements such as walking and changing postures, therefore multitasking becomes more challenging.[22,23] Undertaking multiple tasks simultaneously will create a 'split' in attention leaving some tasks with less focus than others.[36,37] It is therefore understandable that the increase in attentional demand of a task could lead to losing balance and falling due to that shift in focus. Divested attention training could therefore arguably benefit older adults to help train and prepare their sensory, motor and cognitive processes for these situations, helping reduce their falls risk.[24–27]

Limitations

A number of limitations of this study are worth noting, The strategies adopted during the mBESS could have affected scoring. For example, during the one-legged stance one individual might place their foot on the ground and take their time to regain balance before lifting their foot again, whereas another might touch the floor and raise their foot multiple times in attempting to regain balance. Although both techniques could take the same amount of time and have the same balance outcome, they could be scored differently. For the purposes of this study an error was counted as a continuous error until the participant regained control of their balance. Although this helped solve the issue of varying techniques, it raised another issue of the rater's perceptions of when control was regained. As the inter-rater reliability was high it seemed that all raters had similar interpretations of error scoring.

Individual urgency to regain balance also differed noticeably. For older adults, the speed that they are able to move can vary. During testing, individuals perceived as fitter appeared to regain control promptly and therefore had more time in the balance position, whereas slower responders tended to have less time in balance. The conundrum is that more time in balance increases the opportunity to lose balance and be scored an error.

Implications

Current balance tests could be critiqued for having poor external validity due to being overly controlled. There is an identified need for more functionally relevant tests of balance. Movements use more attentional demand as individuals age and training and testing should reflect these requirements. Instead of training that permits the participant's sole focus to be on their balance, there is arguably value in adding a multitasking element [22,23] to balance training and testing. A test such as the mBESS test could be ideal for testing the falls risk of older adults, due to the dual-task element. The next logical step would be a pilot study of the mBESS before using it in clinical practice.

Conclusions

Reducing the incidence and harm of falls would benefit society. This research found that the mBESS test has strong inter-rater reliability, with an average inter-rater reliability of 0.972 across all stances. Due to the low errors and low variability observed for this small cohort, it was hard to judge whether the mBESS was reliable for measuring an individual's balance ability. Ideally further research will explore its reliability and applicability in the real world. Further research needs to be conducted as the idea of divested attention training and assessments could be an important step in preventing falls among older adults.

References

1. Manlapaz DG. Exergaming to improve balance and decrease the risk of falling in individuals with knee osteoarthritis. 2018.
2. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World J*. 2001;1:14. doi:10.1100/tsw.2001.58
3. WHO. World report on ageing and health. *World Heal Organ*. 2015.
4. World Health Organization. Falls. WHO. https://www.who.int/news-room/fact-sheets/detail/falls. Published 2018.
5. Lawler D. Falls data. ACC. Retrieved 4 August, 2020.
6. Al Aama T. Falls in the elderly: Spectrum and prevention. *Can Fam Physician*. 2011;57(7):771-776.
7. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off." *J Am Med Assoc*. 2010;303(3):258-266. doi:10.1001/jama.2009.2024
8. Cameron ID, Gillespie LD, Robertson MC, et al. Interventions for preventing falls in older people in care facilities and hospitals. *Cochrane Database Syst Rev*. 2018;9. doi:10.1002/14651858.CD005465.pub3
9. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012. doi:10.1002/14651858.CD007146.pub3
10. Karlsson MK, Magnusson H, von Schewelov T, Rosengren BE. Prevention of falls in the elderly: a review. *Osteoporos Int*. 2013;24(3):747-762. doi:10.1007/s00198-012-2256-7
11. Kannus P, Sievänen H, Palvanen M, Järvinen T, Parkkari J. Prevention of falls and consequent injuries in elderly people. *Lancet*. 2005;366(9500):1885-1893. doi:10.1016/S0140-6736(05)67604-0
12. Borowicz A, Zasadzka E, Gaczkowska A, Gawłowska O, Pawlaczyk M. Assessing gait and balance impairment in elderly residents of nursing homes. *J Phys Ther Sci*. 2016;28(9):2486-2490. doi:10.1589/jpts.28.2486
13. Health Quality & Safety Commission. An integrated approach to falls in older people: what is your part? 2017:5.
14. Unver B, Kalkan S, Yuksel E, Kahraman T, Karatosun V. Reliability of the 50-foot walk test and 30-sec Chair Stand Test in total knee arthroplasty. *Acta Ortop Bras*. 2015;23(4):184-187. doi:10.1590/1413-78522015230401018
15. Macfarlane DJ, Chou KL, Cheng YH, Chi I. Validity and normative data for thirty-second chair stand test in elderly community-dwelling Hong Kong Chinese. *Am J Hum Biol*. 2006;18(3):418-421. doi:10.1002/ajhb.20503
16. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. *Sports Health*. 2011;3(3):287-295. doi:10.1177/1941738111403122
17. Kleffelgård I, Langhammer B, Sandhaug M, Lundgaard Soberg H. Reliability and validity of the balance error scoring system – BESS. *Physiotherapy*. 2015;101(May):e764-e765. doi:10.1016/j.physio.2015.03.3636
18. NICE. NICE clinical guideline 161. Falls: Assessment and prevention of falls in older people. *Natl Inst Heal Care Excell*. 2013. doi:10.7748/nop.26.6.18.e586
19. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. *BMC Geriatr*. 2014;14(1):14. doi:10.1186/1471-2318-14-14
20. Drootin M. Summary of the updated american geriatrics society/british geriatrics society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc*. 2011;59(1):148-157. doi:10.1111/j.1532-5415.2010.03234.x
21. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler C. Timed up and go test and risk of falls in older adults: A systematic review. *J Nutr Heal Aging*.

2011;15(10):933-938. doi:10.1007/s12603-011-0062-0
22. Mulder T, Zijlstra W, Geurts A. Assessment of motor recovery and decline. *Gait Posture*. 2002;16(2):198-210. doi:10.1016/S0966-6362(01)00157-6
23. Swanenburg J, de Bruin ED, Stauffacher M, Mulder T, Uebelhart D. Effects of exercise and nutrition on postural balance and risk of falling in elderly people with decreased bone mineral density: randomized controlled trial pilot study. *Clin Rehabil*. 2007;21(6):523-534. doi:10.1177/0269215507075206
24. Swanenburg Favero, Uebelhart & Mulder de B. The reliability of postural balance measures in single and dual tasking in elderly falls and non-fallers. *BMC Musculoskelet Disord*. 2008;9(1). doi:10.1186/1471-2474-9-162
25. de Visser E, Veth RP, Schreuder HW, Duysens J, Mulder T. Reorganization of gait after limb-saving surgery of the lower limb. *Am J Phys Med Rehabil*. 2003;82(11):825-831. doi:10.1097/01.PHM.0000091981.41025.FC
26. Scherder E, Eggermont L, Swaab D, et al. Gait in ageing and associated dementias; its relationship with cognition. *Neurosci Biobehav Rev*. 2007;31(4):485-497. doi:https://doi.org/10.1016/j.neubiorev.2006.11.007
27. Dault MC, Geurts ACH, Mulder TW, Duysens J. Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations. *Gait Posture*. 2001;14(3):248-255. doi:10.1016/S0966-6362(01)00130-8
28. Silsupadol P, Shumway-Cook A, Lugade V, et al. Effects of Single-Task Versus Dual-Task Training on Balance Performance in Older Adults: A Double-Blind, Randomized Controlled Trial. *Arch Phys Med Rehabil*. 2009;90(3):381-387. doi:10.1016/j.apmr.2008.09.559
29. Dault MC, Geurts ACH, Mulder TW, Duysens J. Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations. *Gait Posture*. 2001;14(3):248-255. doi:https://doi.org/10.1016/S0966-6362(01)00130-8
30. Agmon M, Belza B, Nguyen HQ, Logsdon RG, Kelly VE. A systematic review of interventions conducted in clinical or community settings to improve dual-task postural control in older adults. *Clin Interv Aging*. 2014. doi:10.2147/CIA.S54978
31. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*.; 2010.
32. University of North Carolina. Balance Error Scoring System (BESS). 2010:1-5.
33. Cicchetti D V. Choice of Agreement Statistics: A Discussion of the Underlying Biostatistics, With Heuristic Examples From the Vineland™–3. *NCS Pearson*. 2017.
34. Cicchetti D V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol Assess*. 1994;6(4):284-290. doi:10.1037/1040-3590.6.4.284
35. Finnoff JT, Peterson VJ, Hollman JH, Smith J. Intrarater and Interrater Reliability of the Balance Error Scoring System (BESS). *PM R*. 2009;1(1):50-54. doi:10.1016/j.pmrj.2008.06.002
36. Wickens C. *Processing Resources in Attention: Varieties of Attention*.; 1991.
37. Young MS, Stanton NA. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Hum Factors*. 2002;44(3):2002. doi:10.1518/0018720024497709

Appendix 2: Scoring sheets

Participant one

Test	Position	Trial	Score
1	1	1	
1	1	2	
1	1	3	
1	2	1	
1	2	2	
1	2	3	
1	3	1	
1	3	2	
1	3	3	
2	1	1	
2	1	2	
2	1	3	
2	2	1	
2	2	2	
2	2	3	
2	3	1	
2	3	2	
2	3	3	
3	1	1	
3	1	2	
3	1	3	
3	2	1	
3	2	2	
3	2	3	
3	3	1	
3	3	2	
3	3	3	

Participant two

Test	Position	Trial	Score
1	1	1	
1	1	2	
1	1	3	
1	2	1	
1	2	2	
1	2	3	
1	3	1	
1	3	2	
1	3	3	
2	1	1	
2	1	2	
2	1	3	
2	2	1	
2	2	2	
2	2	3	
2	3	1	
2	3	2	
2	3	3	
3	1	1	
3	1	2	
3	1	3	
3	2	1	
3	2	2	
3	2	3	
3	3	1	
3	3	2	
3	3	3	

Participant three

Test	Position	Trial	Score
1	1	1	
1	1	2	
1	1	3	
1	2	1	
1	2	2	
1	2	3	
1	3	1	
1	3	2	
1	3	3	
2	1	1	
2	1	2	
2	1	3	
2	2	1	
2	2	2	
2	2	3	
2	3	1	
2	3	2	
2	3	3	
3	1	1	
3	1	2	
3	1	3	
3	2	1	
3	2	2	
3	2	3	
3	3	1	
3	3	2	
3	3	3	

Participant four

Test	Position	Trial	Score
1	1	1	
1	1	2	
1	1	3	
1	2	1	
1	2	2	
1	2	3	
1	3	1	
1	3	2	
1	3	3	
2	1	1	
2	1	2	
2	1	3	
2	2	1	
2	2	2	
2	2	3	
2	3	1	
2	3	2	
2	3	3	
3	1	1	
3	1	2	
3	1	3	
3	2	1	
3	2	2	
3	2	3	
3	3	1	
3	3	2	
3	3	3	

Appendix 3: Scorer instructions

Scorer Instructions

You will be scoring tests derived from the Balance Error Scoring System.

Each test you will watch is 20 seconds long. During this time it is your job to identify as many errors that occur as possible. This score can be anything from 0+.

The three stances performed by the participants are:

- Parallel stance (feet side by side, heels touching)
- Tandem stance (heel toe, with dominant leg behind)
- One-legged stance (standing on one leg)

An error is scored when a participant:

- Lifts their hands off their hips
- Opens their eyes
- Steps, stumbles or falls
- Hip abduction of 30 degrees or more
- Lifting their forefoot or heel
- Remaining out of position for 5+ seconds.

Two errors at the same time counts for only one error on the score card.

I recommend quickly working out what 30 degrees would look like before starting scoring. As we are trying to emulate what this would be like for someone in real-time please try and only watch each test once and score.

The timer is on the screen to show you when to start and stop counting errors.

You may find:

- Position one will have many zero scores due to no errors.
- Some participants will lift their hands to touch their face – please DO NOT count this as an error, please only count their hands lifting if it's to help with balance.

What to expect:

Each participant did 3 testing sessions, each session they performed 3 trials in parallel stance, 3 trials in tandem stance and 3 trials in one-legged stance. In the recording of the participants each test and trial will be shown in order of filming. An Excel spreadsheet is attached for you to fill in the scores. Each participant has a scorecard which you can fill in.

Appendix 4: Ethical approval

27 January 2020

Lily Purdon
c/- Institute of Sport, Exercise and Health
Otago Polytechnic
Private Bag 1910
Dunedin 9054

Dear Lily

Ethics approval for project
Reference Number: 837
Application Title: *Is the Balance Error Scoring System test in combination with a distractive task a reliable measure of balance ability?*

Thank you for your application for ethics approval for this research project. This letter is to advise that the Otago Polytechnic Research Ethics Committee review panel has approved your application, following amendments made in response to feedback; and also **conditional** on the following further corrections being made:

Recruitment Poster
1. Change email address – it is still incorrect
2. Correct typos/wording: "Participant involves" to Participation involves. "Help better the balance of elderly". Change to Help better understand the balance of elderly or better assist elderly with balance.
3. Change "someone whose been affected by lose of balance or falling" to someone who has been affected by loss of balance or falling.
4. Correct address. Logan Park Drive is in North Dunedin.

Participant Information Sheet - Change the line at bottom to: "This project has been approved by the Otago Polytechnic Research Ethics Committee".

We wish you well with your work and remind you that at the conclusion of your research to send a brief report with findings and/or conclusions to the Ethics Committee. All correspondence regarding this application should include the project title and reference number assigned to it.

This protocol covers the following researchers: Lily Purdon, Phil Handcock and Richard Humphrey.

Regards

Dr. Emilie Crossley
Otago Polytechnic Research Ethics Committee