# Reliability and concurrent validity of the iPhone® Compass application to measure thoracic rotation range of motion (ROM) in healthy participants

**James Furness** [Corresp., 1, 2] , **Ben Schram** [1, 2] , **Alistair J Cox** [2] , **Sarah L Anderson** [2] , **Justin Keogh** [3, 4, 5]

[1] Water Based Research Unit, Bond Institute of Health and Sport, Bond University, Gold Coast, Queensland, Australia
[2] Department of Physiotherapy, Faculty of Health Science and Medicine, Bond University, Gold Coast, Queensland, Australia
[3] Faculty of Health Science and Medicine, Bond University, Gold Coast, Queensland, Australia
[4] Sports Performance Research Centre New Zealand, AUT University, Auckland, New Zealand
[5] Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Sunshine Coast, Australia

Corresponding Author: James Furness
Email address: jfurness@bond.edu.au

**Background:** Several water based sports (swimming, surfing and stand up paddle boarding) require adequate thoracic mobility (specifically rotation) in order to perform the appropriate activity requirements. The measurement of thoracic spine rotation is problematic for clinicians due to a lack of convenient and reliable measurement techniques. More recently, smartphones have been used to quantify movement in various joints in the body, however there appears to be a paucity of research using smartphones to assess thoracic spine movement. Therefore, the aim of this study is to determine the reliability (intra and inter rater) and validity of the iPhone® app (Compass) when assessing thoracic spine rotation ROM in healthy individuals. **Methods** A total of thirty participants were recruited for this study. Thoracic spine rotation ROM was measured using both the current clinical gold standard, a universal goniometer (UG) and the Smart Phone Compass app. Intra-rater and inter-rater reliability was determined with a Intraclass Correlation Coefficient (ICC) and associated 95% confidence intervals (CI). Validation of the Compass app in comparison to the UG was measured using Pearson's correlation coefficient and levels of agreement were identified with Bland-Altman plots and 95% limits of agreement. **Results:** Both the UG and Compass app measurements both had excellent reproducibility for intra-rater (ICC 0.94-0.98) and inter-rater reliability (ICC 0.72-0.89). However, the Compass app measurements had higher intra-rater reliability (ICC = 0.96-0.98; 95% CI 0.93-0.99; vs. ICC = 0.94-0.98; 95% CI 0.88-0.99) and inter-rater reliability (ICC = 0.87-0.89; 95% CI 0.74-0.95 vs. ICC = 0.72-0.82; 95% CI 0.21-0.94). A strong and significant correlation was found between the UG and the Compass app, demonstrating good concurrent validity (r =0.835, p<0.001). Levels of agreement between the two devices were 24.8° (LoA –9.5°, +15.3°). The UG was found to consistently measure higher values

than the compass app (mean difference 2.8°, P<0.001). **Conclusion:** This study reveals that the iPhone® app (Compass) is a reliable tool for measuring thoracic spine rotation which produces greater reproducibility of measurements both within and between raters than a UG. As a significant positive correlation exists between the Compass app and UG, this supports the use of either device in clinical practice as a reliable and valid tool to measure thoracic rotation. Considering the levels of agreement are clinically unacceptable the devices should not be used interchangeably for initial and follow up measurements.

1 **Title:** Reliability and concurrent validity of the iPhone® Compass application to measure thoracic

2 rotation range of motion (ROM) in healthy participants.

3 **Authors:**

4 James Furness (BPhty, PhD) [1,2]

5 Ben Schram (B.Ex.Sci, DPhty., PhD) [1,2]

6 Alistair J. Cox (DPhty) [2]

7 Sarah L. Anderson (DPhty) [2]

8 Justin Keogh (BHMS(Hons), BHSc, PhD) [3,4,5]

9     *1. Water Based Research Unit, Bond Institute of Health and Sport, Bond University, Gold*
10        *Coast, QLD, Australia.*
11     *2. Department of Physiotherapy Faculty of Health Science and Medicine, Bond University,*
12        *Gold Coast, QLD, Australia.*
13     *3. Faculty of Health Science and Medicine, Bond University, Gold Coast, QLD, Australia*
14     *4.  Sports Performance Research Centre New Zealand, AUT University, Auckland, New*
15        *Zealand*
16     *5.  Faculty of Science, Health, Education and Engineering, University of the Sunshine*
17        *Coast, Australia*

18 **Corresponding Author:**

19 James Furness

20 Bond Institute of Health and Sport, 2 Promethean Way, Robina, QLD, Australia.

21 P: (07) 5595 3354

22 E: jfurness@bond.edu.au

23    **ABSTRACT**

24    **Background:** Several water based sports (swimming, surfing and stand up paddle boarding)

25    require adequate thoracic mobility (specifically rotation) in order to perform the appropriate

26    activity requirements. The measurement of thoracic spine rotation is problematic for clinicians

27    due to a lack of convenient and reliable measurement techniques. More recently, smartphones

28    have been used to quantify movement in various joints in the body, however there appears to be a

29    paucity of research using smartphones to assess thoracic spine movement. Therefore, the aim of

30    this study is to determine the reliability (intra and inter rater) and validity of the iPhone® app

31    (Compass) when assessing thoracic spine rotation ROM in healthy individuals.

32    **Methods** A total of thirty participants were recruited for this study. Thoracic spine rotation ROM

33    was measured using both the current clinical gold standard, a universal goniometer (UG) and the

34    Smart Phone Compass app. Intra-rater and inter-rater reliability was determined with a Intraclass

35    Correlation Coefficient (ICC) and associated 95% confidence intervals (CI). Validation of the

36    Compass app in comparison to the UG was measured using Pearson's correlation coefficient and

37    levels of agreement were identified with Bland-Altman plots and 95% limits of agreement.

38    **Results:** Both the UG and Compass app measurements both had excellent reproducibility for

39    intra-rater (ICC 0.94-0.98) and inter-rater reliability (ICC 0.72-0.89). However, the Compass app

40    measurements had higher intra-rater reliability (ICC = 0.96-0.98; 95% CI 0.93-0.99; vs. ICC =

41    0.94-0.98; 95% CI 0.88-0.99) and inter-rater reliability (ICC = 0.87-0.89; 95% CI 0.74-0.95 vs.

42    ICC = 0.72-0.82; 95% CI 0.21-0.94). A strong and significant correlation was found between the

43    UG and the Compass app, demonstrating good concurrent validity (r =0.835, p<0.001). Levels of

44    agreement between the two devices were 24.8° (LoA −9.5°, +15.3°). The UG was found to

45    consistently measure higher values than the compass app (mean difference 2.8°, P<0.001).

46    **Conclusion:** This study reveals that the iPhone® app (Compass) is a reliable tool for measuring

47    thoracic spine rotation which produces greater reproducibility of measurements both within and

48    between raters than a UG. As a significant positive correlation exists between the Compass app

49    and UG, this supports the use of either device in clinical practice as a reliable and valid tool to

50    measure thoracic rotation. Considering the levels of agreement are clinically unacceptable the

51    devices should not be used interchangeably for initial and follow up measurements.

## INTRODUCTION

52

53  Thoracic spine mobility is a key component in many water-based sports including surfing,

54  swimming and stand up paddleboarding (SUP) (Furness 2015; Pollard & Fernandez 2004;

55  Schram 2015). In surfing, thoracic rotation is a critical movement during wave riding as it allows

56  the surfer to produce sufficient torque to turn and perform maneuvers (Furness et al. 2016). It is

57  also hypothesised that a reduction in thoracic extension during the paddling phase of surfing

58  could potentially result in greater stress on the lumbar or cervical spine (Furness et al. 2014).

59  Thoracic rotation is required during the pull phases of freestyle and backstroke when swimming

60  and forms a crucial component of the kinetic chain during stroke initiation as the athlete attempts

61  to reach a position of high humeral elevation (Blanch 2004). Finally in SUP, efficient paddling

62  occurs when the trunk is rotating to transfer force through the kinetic chain and propel the

63  paddler past the blade (Schram 2015).

64  Limitations to range of motion (ROM) within the thoracic spine can create compensatory

65  mechanisms that require a greater motion requirement from the scapulothoracic and

66  glenohumeral joints leading to muscle and strength asymmetries, and increased injury risk

67  (Blanch 2004; Spurrier 2006). Given the importance of thoracic spine mobility to this cohort, an

68  assessment of rotation would be a logical inclusion in any musculoskeletal assessment of water

69  based athletes. The purpose of such an assessment is to develop a profile of the athlete relating to

70  their past and current injuries and identify limitations to joint ROM. This would allow the

71  therapist to select interventions specific to the athlete's musculoskeletal deficits which reduces

72  injury risk and enhances athletic performance (Furness 2015; Spurrier 2006). Common

73  musculoskeletal pathologies involving the cervical spine, shoulder and lumbar spine have been

74  treated successfully by identifying inconsistencies in strength and or ROM of the thoracic spine

75  (Iveson et al. 2010). The ability to quantify these thoracic asymmetries through reliable and

76    feasible methods are limited in the literature making it difficult to comprehensively assess the

77    thoracic spine (Iveson et al. 2010).

78    In current clinical practice, a goniometer is commonly used by physiotherapists due to its ease of

79    use, cost effectiveness, and reliability, generally ranging from good to excellent depending on the

80    extremity being assessed (Norkin & White 2016). Studies of reliability of the measurement of

81    spinal movements using goniometry are limited however, with such measures reported to have

82    lower reliability than those of the extremities (Burdett et al. 1986; Gajdosik & Bohannon 1987;

83    Nitschke et al. 1999; Youdas et al. 1991). This may be due to the complexity of the spine and

84    position of surrounding tissues which can make it difficult to palpate anatomical landmarks to use

85    as a reference points for goniometer positioning. These characteristics of the spine also can lead

86    to movement of the axis or stationary arm of the goniometer, predisposing it to error. Both the

87    individuals effort in active ROM and variations in manual force applied by the examiner in

88    passive ROM can also be a source error in goniometry (Norkin & White 2016).

89    A novel solution to negating the potential errors inherent to goniometry is in the use of a smart

90    phone application. More recently, smartphones have been used to quantify joint movement and

91    may offer equal or greater reliability than goniometry while avoiding several of the sources of

92    error. Smartphone software applications (apps) programmed to measure joint ROM may provide

93    health practitioners a novel clinical tool to examine joint function, detect joint asymmetry and

94    evaluate treatment efficacy as an objective outcome measure. They are easy to use, cheap and

95    highly accessible, with an estimated 79% of Australians under 55 -year-olds owning a smart

96    phone (Drumm & Swiegers 2015). A  recent systematic review evaluating mobile phone based

97    tools to assess joint ROM, functional assessment and rehabilitation of proprioception identified

98    22 articles which measured either reliability and or validity (Mourcou et al. 2015). Several of

99    these articles used clinical tools as the gold standard reference (goniometer, inclinometer,

100   scoliometer etc.) (Jenny et al. 2016; Jones et al. 2014; Tousignant-Laflamme et al. 2013),

101   however no study was identified that  utilized a mobile phone application to  determine thoracic

102   spine rotation ROM (reliability or Validity).

103   Given the importance of thoracic rotation in water based sports, simple clinical measures to

104   quantify rotation need to be developed. Despite being used successfully in various other joints in

105   the body, smartphone app assessment of thoracic spine rotation is yet to be investigated.

106   Therefore, the aim of this study is to determine the reliability (intra and inter rater) and validity of

107   the iPhone® app (Compass) when assessing thoracic spine rotation ROM in healthy individuals.

108  **METHODS**

109  *Study Design*

110  Observational design assessing the reliability and concurrent validity of the iPhone® app

111  (Compass).

112  *Participants*

113  A total of thirty participants were recruited from a student University population. Sample sizes of

114  at least 15-20 is considered adequate for reliability studies which collect continuous data (Lexell

115  & Downham 2005). Participants were excluded if they were currently experiencing back or trunk

116  pain, had any back injury within 6 weeks prior to testing, had a history of spinal surgery, were

117  younger than 18 years of age, or refused to give informed consent to perform thoracic rotation

118  from a seated position. The study was approved by the University Human Research Ethics

119  committee (RO 1610) and informed consent was gained from all participants via a consent form.

120  *Examiners*

121  Measurements were conducted by two final year post graduate physiotherapy students. Both

122  students underwent several familiarization sessions using the UG and the Compass app. These

123  sessions  were supervised by two senior physiotherapists with over five years of experience in

124  assessment of orthopedic conditions. In addition, each examiner familiarized themselves with

125  both the Compass app and the measurement of thoracic rotation prior to any measurements. They

126  were blinded to each other's results by independently recording data into separate spreadsheets.

127  *Instrumentation*

128   For this study, a UG and an iPhone® model 6S (iPhone® is a trademark of Apple Inc, Cupertino,

129   California, USA) with the Compass app (Apple Pty Limited, United States) were  utilised. The

130   UG was a plastic twelve-inch goniometer (SAEHAN Grip™ Rulong 20cm, Belgium) with

131   moveable arms and a numerical face. The Compass app (Apple Pty Limited) comes pre-installed

132   on all iPhones and uses the device's inbuilt magnetometer, accelerometer and GPS receiver to

133   detect the Earth's magnetic field and orients the device to magnetic north while the device rotates

134   through the transverse plane (Dixon-Warren 2012; Wilson & Fenlon 2007). It has been validated

135   for recording ROM measurements in the transverse plane in the cervical spine (Tousignant-

136   Laflamme et al. 2013). Both examiners used an Apple iPhone® 6S (64GB, model no.

137   NKQR2ZP/A, Cupertino, CA, USA) running iOS 10.2.

138   *Experimental Procedures*

139   For this study thoracic rotation was defined as the combined gross axial rotation occurring

140   throughout all thoracic spinal segments. As the shoulders, cervical and lumbar spine and hips all

141   contribute to rotation, several minimization techniques were implemented as discussed below.

142   There are several techniques reported in the literature for measuring thoracic rotation, however

143   the 'seated rotation with bar in front' technique has been reported to produce the greatest within-

144   day inter-rater reliability (ICC: 0.87) and was therefore utilized for this study (Johnson et al.

145   2012; Norkin & White 2016). This technique allows easier visualization of anatomical landmarks

146   making it more suitable for novice examiners (Johnson et al. 2012).

147   The order of examiners, UG or Compass app and the direction in which the participant rotated

148   were randomized using the randomization function in Microsoft Excel (v16.0, Washington,

149   USA). After one examiner had completed all measurements, participants were asked to return

150   five minutes later so that the next examiner could repeat the same protocol. This timeframe has

151    previously been used in a reliability study assessing thoracic rotation (Johnson et al. 2012). The

152    examiners obtained  three active ROM measurements for left and right rotation for both the

153    Compass App and the Goniometer. This provided measures for within-day intra and inter-rater

154    reliability. All measured angles were recorded into an Excel spreadsheet by the examiner

155    immediately after the measurement was taken. Each examiner had their own spreadsheet to

156    ensure true blinding between examiners.

157    Prior to any measurements being taken, subjects were asked to complete a series of warm up

158    exercises to ensure task familiarity and to minimise the chance of a warmup effect on ROM or

159    injury. This warm up involved repeated submaximal ROM thoracic rotations to the left and right

160    five times whilst seated with their arms across their chest as per the ACSM guidelines for

161    movement specificity (Nadalen 2016). To minimize variations in participant positioning,

162    participants were instructed to place their feet flat on the floor, knees and hips in 90° of flexion

163    and to maintain a neutral spine position where the lumbar lordosis was maintained. To minimize

164    contribution of movement in the hips, an inflatable ball was placed between the participant's

165    knees. To minimize contribution of movement in the upper limbs primarily scapula-thoracic

166    region, participants crossed their arms over their chest while holding a wooden bar which was

167    placed at the acromioclavicular level (Fig 1) (Johnson et al. 2012). The examiners provided

168    standardized verbal (See Appendix A) and manual cues to correct any deviations from this

169    positioning throughout testing to ensure consistency of the movement performed with each trial.

170    **Insert Figure 1 here 'Starting position for seated rotation bar-in-front technique'.**

171    *Universal Goniometer Measurement,*

172    The UG was positioned with the axis fixed in the transverse plane at the level of T1-T2. As

173    previously used by Lewis & Valentine (2010), T1-T2 was identified by palpating inferiorly from

174    C7 and by palpating the spine of the scapula and moving medially to the spinous process and

175    moving up 1-2 levels. The stationary arm of the UG was pointed to the opposite side of rotation

176    and remained in the starting position while the mobile arm remained parallel to the floor, with the

177    spine of the scapula used a reference point during rotation (See Fig 2). Participants were asked to

178    look straight ahead to minimise any cervical rotation. The maximum angle the participant reached

179    at the end of their ROM was both measured and recorded by the same examiner. At the

180    completion of measurement, the participant was instructed to return to the start position where the

181    UG was removed from the subject and repositioned before the next measurement. Participants

182    were allowed to rest while the examiner entered the measurement into a spreadsheet.

183    **Insert Figure 2 here 'Measurement of thoracic rotation with the universal goniometer'.**

184    *Compass App Measurement*

185    To measure thoracic rotation with the Compass app, the participant was placed in the position

186    shown in Fig. 1. The examiner placed the iPhone® approximately at the level of T1-T2 of each

187    participant using the same palpation method described previously. The iPhone® was positioned so

188    that magnetic north was facing directly towards the participant which was determined by the

189    Compass app dial reading 0°. The examiner held the iPhone® firmly against the participants back

190    at the T1-T2 level whilst the participant performed active thoracic rotation (Figure 3) to their end

191    ROM. The examiner followed the participant's movement by maintaining firm pressure of the

192    iPhone® against the T1-T2 level during rotation. The participant was asked to return to the

193    starting position which was determined by the Compass app displaying  a value of 0° position

194  before the subsequent measurement; this was actively completed by the participant. Participants

195  were allowed a period of rest while the examiner entered each measure into a spreadsheet.

196  **Insert Figure 3 here 'Measurement of thoracic rotation with the iPhone Compass App'.**

197  ***Statistical Analysis***

198  Statistical analyses were performed using SPSS (Version 24.0; IBM Corp, Armonk, NY). Intra-

199  rater and inter-rater reliability  were determined using an intraclass correlation coefficient ($ICC_{3.2}$

200  $_{and}$ $ICC_{2.2}$ respectively) and associated 95% confidence intervals (CIs) (Trevethan 2017). Shrout &

201  Fleiss (1979) suggest that ICC values greater than 0.75 indicate excellent reproducibility, values

202  between 0.40 and 0.75 represent fair to good reliability and values less than 0.40 indicate poor

203  reproducibility.

204  To assess measurement variability, the standard error of measurement (SEM) was calculated. This

205  measurement is defined by the equation $SEM = \sqrt{WMS}$ whereby WMS represents the mean

206  square error term from the analysis of variance (Lexell & Downham 2005). Furthermore, the

207  SEM provides an 'absolute index of reliability' or 'typical error' associated with a measurement

208  (Trevethan 2017). The smallest amount of change that can be detected by a measure that

209  corresponds to a noticeable change (clinically important changes) was calculated using the

210  smallest real difference (SRD) equation $SRD = 1.96 \times SEM \times \sqrt{2}$.

211  To determine the construct validity of the Compass app, a Pearson's correlation coefficient was

212  used. The linear relationship was presented graphically in a scatter plot with the associated R

213  squared value. The level of agreement between the two devices was also presented through Bland

214    Altman plots with the associated 95% limits of agreement. In this, the formula: mean difference

215    between measures ±1.96 x SD, was used (Bland & Altman 2010).

216    **RESULTS**

217    *Demographics*

218    In total, 30 participants were assessed for their thoracic rotation (20 females, 10 males, 29.8 ±8.9

219    years, height 167.8±8.9cm, mass 67.9±10.1kg. Average range of motion measurements are

220    displayed in Table 1. Measurement of thoracic rotation with the UG resulted in an average score

221    of 63.0° (±11.3) while the Compass app averaged 60.1° (±10.7). Paired sample t-tests revealed no

222    differences between left and right rotation for both the Goniometer and Compass App (p = 0.279

223    and p = 0.791 respectively).

224    **Insert Table 1 here '**Descriptive Data for Thoracic Rotation Values'

225    *Reliability Analysis*

226    *Intra-rater reliability*

227    Examiners 1 and 2 had consistently high ICC values for both the UG and the Compass app as

228    shown in Table 2. ICC values for examiner 1 were between 0.97-0.98 and between 0.94-0.97 for

229    examiner 2 for both instruments indicating excellent reproducibility.

230    **Insert Table 2 here 'Intra-rater reliability for examiner 1 and 2 comparing the UG and the iPhone® Compass**

231    **application'**.

232    *Inter-rater reliability*

233    ICC values for the UG and the Compass app for both examiners were in the good-excellent range

234    (0.72-0.89) (Table 3). The Compass app showed slightly higher ICC values than the UG, with

235    less variability as indicated by the upper and lower confidence intervals.

236    **Insert Table 3 here 'Inter-rater reliability for the UG and the iPhone® Compass application'.**

237    *Concurrent Validity*

238    Pearson's correlation coefficient revealed a strong and significant association ($r = 0.835$,

239    P<0.001) between the UG and the Compass App measurements. As no significant differences

240    were seen between left and right rotation for either device an average of both movements was

241    used in the analysis. Figure 4 presents this positive linear association between the two devices.

242    Linear regression analysis was performed which resulted in a value of $r^2 = 0.697$, indicating that

243    approximately 70% of the variation in values obtained by the compass app can be explained by

244    the variation of values obtained by the UG. .

245    A Bland-Altman plot was produced to graphically represent the level of agreement between the

246    two devices for thoracic rotation (left and right averaged) (Fig 5). The mean difference between

247    measures was 2.8° (SD 6.3°) and the upper and lower limits of agreements were 15.3° and –9.5°,

248    respectively. Figure 5 graphically demonstrates the majority of data points close to the mean

249    difference and within the 95% limits of agreement; however, values can vary by 24.8°.

250    Systematic bias was revealed through a one sample t test (2.8°, P<0.001), indicating UG values

251    were consistently higher when compared with the Compass App (the null hypothesis would result

252    in a mean difference of 0°).

253  **Insert Figure 4 here 'Scatterplot Showing Relationship between iPhone® Compass app and Goniometer for**
254  **thoracic rotation ROM'.**

255  **Insert Figure 5 here 'Bland–Altman Plot indicating mean difference and 95% limits of agreement between**
256  **measurements from the iPhone® Compass app and Goniometer for thoracic rotation (°)'.**

257  **DISCUSSION**

258  The aim of this study was to determine the reliability (intra and inter rater) and validity of the

259  iPhone® app (Compass) when assessing thoracic spine rotation ROM in healthy individuals. The

260  primary findings of this study reveal that both the UG and Compass app measurements

261  demonstrated excellent reliability; however, ICC values for the Compass app were greater for

262  both intra-rater and inter-rater reliability. Finally, a strong and significant correlation was also

263  shown between the devices.

264  The mean values for thoracic rotation ROM were 60.1° (±10.7) with the Compass app tended to

265  be approximately 2-3° higher at 63.0° (±11.3) with the UG (Table 1). These values were similar

266  to previous studies which report average thoracic rotation ROM to be 55.4° (±9.2) when using

267  the same technique (Johnson et al. 2012). The consistently higher values recorded by the UG are

268  thought to be related to its positioning during measurement. The moveable arm is held against the

269  posterior surface of the shoulder and is therefore exposed to the movement of the scapula which

270  often tilts anteriorly as participants reach their end ROM. It is hypothesized that due to the

271  iPhone's position being localised to the T1-T2 level, its movement is not influenced by the

272  contribution of the scapulothoracic joints during spine rotation. Considering the above the authors

273  believe the iPhone application provides a rotation recording which is more localised to spinal

274  ROM than that of the UG recording'.

275   *Intra-rater reliability*

276   An excellent intra-rater reliability was found when either the Compass app (ICC = 0.96-0.98;

277   95% CI 0.93-0.99) or UG (ICC = 0.94-0.98; 95% CI 0.88-0.99) was used to assess thoracic

278   rotation ROM. While both instruments demonstrated excellent intra-rater reliability, the ICC

279   values are marginally higher for the Compass app. Moreover, 95% CI ranges were shown to be

280   wider for the UG compared to the Compass app, indicating greater variability when the UG is

281   used to measure thoracic rotation. In agreement with the present findings, Johnson et al. (2012)

282   reported excellent within-day intra-rater reliability for UG measurement of thoracic rotation using

283   the same technique, however, ICC values were slightly lower and 95% CI ranges were wider

284   (ICC = 0.87-0.91; 95% CI 0.76-0.95). While these differences are only small, they could be

285   explained by the differences between methodologies. The measurement technique in the current

286   study used verbal cues with each measurement such as "rotate as far to the left/right as you

287   comfortably can". The purpose of this was to reduce variations in the individual's effort during

288   active ROM which is known to be a common source of error in goniometry (Norkin & White

289   2016). Johnson et al. (2012) did not report use of verbal cues in their study which may have

290   resulted in lower ICC values and wider 95% CI ranges. Tousignant-Laflamme et al. (2013) also

291   used the Apple iPhone® app *Compass* to measure transverse plane joint ROM, but in the cervical

292   spine. In contrast to our findings, the authors reported moderate to good intra-rater reliability

293   (ICC = 0.66-0.74; 95% CI: 0.39-0.87) for the Compass app when measuring cervical rotation

294   ROM. Their lower ICC values might be explained by the position of the iPhone® in relation to

295   surrounding structures. In the current study, the iPhone® was placed firmly against the

296   participants back at the T1-T2 level, as opposed to placing the compass on the participant's head

297   in order to measure cervical ROM. This could expose the compass to additional movements

298   outside of rotation about the transverse axis.

299  *Inter-rater reliability*

300  The inter-rater reliability of the present study was excellent when using the Compass app (ICC =

301  0.87-0.89; 95% CI 0.74-0.95) and good-excellent when using the UG (0.72-0.85; 95% CI 0.21-

302  0.94). The 95% CI ranges were shown to be wider when the UG was used compared to the

303  Compass app. Again, this indicates greater variability of data when the UG is used to measure

304  thoracic rotation ROM. In comparison to our intra-rater reliability findings, both devices had

305  lower ICC values for inter-rater reliability which is a common theme among studies of joint

306  measurement methods (Johnson et al. 2012; Norkin & White 2016). Our ICC values for inter-

307  rater reliability tend to align with those reported in several other studies on smartphone

308  applications and joint ROM measurement. Otter et al. (2015) measured first metatarsalphalangeal

309  joint dorsiflexion with a smartphone goniometer application (Dr. Goniometer) and a UG and

310  reported good inter-rater reliability for the smartphone application (ICC 0.70; 95% CI 0.60-0.80)

311  and moderate-good inter-rater reliability for the UG (ICC 0.69; 95% CI 0.58-0.79). Similarly,

312  Mitchell et al. (2014) reported greater ICC values for inter-rater reliability when comparing two

313  smartphone applications (Dr. Goniometer and GetMyROM) (ICC 0.92-0.94; 95% CI 0.85-0.98)

314  with a UG (ICC 0.91; 95% CI 0.64-0.97) on active shoulder external rotation ROM of 94

315  females. Similar to our study, the CI's reported by Mitchell et al. (2014) were narrower for

316  smartphone applications than UG which again highlights the variability of data associated with

317  UG when compared to the Compass App.

318  *Concurrent validity*

319  A strong, significant correlation was found between the Compass app and UG ($r = 0.835$,

320  P<0.001), demonstrating that measures taken by the Compass app were concurrently valid when

321     compared to the UG. In addition, linear regression analysis revealed a value of $r^2 = 0.697$,

322     indicating that approximately 70% of the variation in values obtained by the compass app can be

323     explained by the variation of values obtained by the UG.. There was a tendency for the UG to

324     consistently measure higher than the Compass app (2.8°, P<0.001, Figure 5), revealing a level of

325     systematic bias between the two measurement approaches.

326     Level of agreement between the devices was explored through Bland Altman plots and

327     calculating the Limits of agreement (LoA). This resulted in a value of 24.8° (LoA –9.5°, +15.3°)

328     highlighting that 95% of differences between measurements by the Compass app and UG will lie

329     within a range of 24.8°.

330     This form of analysis highlights that a high correlation does not mean that the two methods agree

331     (Bland & Altman 2010). Whether the difference between the measures is acceptable comes down

332     to clinical judgement (Bland & Altman 2010). It was decided by the research team that this

333     difference was clinically unacceptable, especially if the clinician wished to use these two devices

334     interchangeably. It was deemed clinically unacceptable as average thoracic ROM tends to be

335     approximately 40-50 degrees (Iveson et al. 2010; Johnson et al. 2012). A difference of 24 degrees

336     would be approximately half the normal thoracic rotation ROM and therefore using either device

337     interchangeably would reveal significant inconsistencies.  The authors recommend consistent

338     choice of measurement instrument which will ultimately improve clinician confidence that any

339     changes to joint ROM are a result of treatment rather than the device used.

340     Considering that this is the first study to assess reliability and validity of the Compass App when

341     measuring Thoracic rotation, it is difficult to draw similarities between our results and previous

342     research. Tousignant-Laflamme et al. (2013) used the Compass app to measure cervical rotation

343     but compared values to the Cervical Range of Motion Device (CROM) as their accepted gold

344   standard. They reported moderate validity for right rotation (ICC 0.55; 95% CI 0.23-0.76) and

345   poor validity for left rotation (ICC 0.43; 95% CI 0.08-0.69). In addition, Pearson correlation

346   values reflected lower validity when compared to our findings (R rotation: $r = 0.58$, P<0.01; L

347   rotation:  $r = 0.38$, P=0.04). The authors concluded that as lower results for validity were only

348   associated with rotation, the Compass app is sensitive to electromagnetic fields in this plane of

349   movement. In the lumbar spine, the smartphone apps TiltMeter© and iHandy© level have both

350   demonstrated high validity ($r \geq 0.86$) when compared to a gravity-based inclinometer for

351   measurement of various standing lumbar spine movements and standing lumbar lordosis (Kolber

352   et al. 2012; Salamh & Kolber 2014). Beyond the spine, the validity of smartphone apps to

353   measure ROM of the joints of the extremities tends to be slightly higher than the results of the

354   present study. Shin et al. (2012) used a smartphone app (Clinometer) to measure various shoulder

355   movements, and when compared to UG measurements, Pearson r values reflected a strong

356   positive correlation between the devices ($r = 0.79$-0.97). Mitchell et al. (2014) examined the

357   smartphone apps GetMyROM, an inclinometry based app, and DrGoniometer, a photo-based app,

358   for assessment of active shoulder ER ROM. The authors reported excellent validity when

359   compared to measures obtained from a UG (GetMyRom: ICC 0.94; 95% CI 0.92–0.96; Dr.

360   Goniometer: ICC 0.93; 95% CI 0.42–0.98). Several other studies have investigated smartphone

361   app measurement of knee joint flexion and have found large positive correlations with UG

362   measures. Both Ockendon & Gilbert et al. (2012) and Jones et al. (2014) reported high Pearson $r$

363   values ($r \geq 0.95$) for the knee goniometer app (KGA) and Simple Goniometer app, respectively.

364   Milanese et al. (2014) also used the KGA app and showed excellent validity when comparing

365   measurements obtained from a UG for knee flexion angle which was reported as concordance

366   correlation coefficients (CCC) (CCC 0.99; 95% CI 0.98-0.99).

367   When comparing the validity data of studies which measured the joints of the extremities with

368   our own results, it is important to consider that the validity and reliability of measurement of joint

369   angles can be adversely affected by the complexity of the joint itself (Gajdosik & Bohannon,

370   1987). The knee, for example, is a uniplanar hinge joint while the thoracic spine is a complex

371   multiplanar, multifaceted joint. In addition, the thoracic spine has multiple adjacent joints and its

372   motion requires contraction of multi-joint musculature. It has been proposed that such

373   characteristics are likely to reduce the reliability and hence validity for measurements of these

374   joints which provides an explanation for why our validity data, while demonstrating a strong

375   positive correlation ($r = 0.835$, P<0.001), was slightly lower in comparison to the majority of

376   studies reviewed (Jones, 2014; Milanese, 2014; Mitchell, 2014; Norkin, 2016; Shin, 2012).

377   *Strengths & Limitations*

378   To authors knowledge the UG has not been validated against a gold standard such as X-ray and

379   Computer Tomography with respect to measuring thoracic rotation ROM even though these

380   methodologies exist (Lam et al. 2008). Given the fact that there appears to be several clinical

381   methods to measure thoracic rotation (tape measure, inclinometer and goniometer) there is no

382   clear consensus as to which tool is superior clinically (Iveson et al. 2010; Johnson et al. 2012;

383   O'Gorman & Jull 1987). Goniometers are the most commonly and easily used tools designed to

384   measure ROM in clinical settings and therefore were selected as the comparative tool to assess

385   the concurrent validity of the Compass app.

386   The measurement of thoracic rotation with an UG is inherently difficult due to the requirement of

387   the stationary arm to remain immobile. There is a tendency for the stationary arm to move a few

388   degrees while manipulating the moving arm. The requirement of the measurer to remain vigilant

389   in observing the stationary arm meant that the measurer couldn't be blinded to their own

390 measurement. This inability to blind the assessor may lead to a degree of measurement bias. The

391 assessors were blinded to each other's measurements however, minimizing recall bias and

392 improving internal validity. The Compass app is also not without limitations which may include;

393 measurements with an iPhone may be subject to radiofrequency noise; secondly the user is

394 unable to calibrate the sensors within the iPhone as this is done at a manufacturing level. This

395 may be more relevant when using older models of the iPhone. To negate this issue, it is advisable

396 to conduct validity/accuracy assessment prior to the commencement of testing by either using a

397 newer model of the iPhone concurrently or by using a geometric tool as a cross reference when

398 measuring angles.

399 The absence of between day measurements may also be viewed as a limitation. It should be noted

400 that clinicians will often perform measurements, treat and then reassess the effects of their

401 treatments however over the course of a single treatment highlighting that the same day reliability

402 is still of clinical value. Future research could investigate the between-day reliability of the

403 Compass app to measure movements in the transverse plane since clinicians will also measure

404 over the course of a few days or weeks during a treatment period.

405 The results of this study are only applicable to the healthy population which was assessed in this

406 study and may not be applicable to people in pathological populations. Future studies should

407 investigate the reliability of the Compass app to measure thoracic spine rotation in participants

408 with thoracic spine injuries or pathologies that are likely to induce limitations to ROM.

409 Strengths of this study include the randomization of the movement direction, examiner and the

410 device. This negated the effect of a warm up and helped to eliminate selection bias thus

411 promoting the efficacy of the protocol (Suresh 2011). The standardization of procedures

412 including instructions to participants, the same plinth, wooden bar, ball, UG and iPhone® was also

413    used for each participant. The authors recommend the same standardization of protocols to be

414    used clinically if this measurement technique is applied in clinical practice.

415    **CONCLUSION**

416    Measurement of thoracic spine rotation ROM has been problematic for physiotherapists due to a

417    lack of convenient and reliable measurement techniques. Both the UG and Compass app for

418    iPhone® offer reliable methods for measurement when the 'seated rotation bar in front' technique

419    is adhered to, however, the intra-rater and especially inter-rater reliability of the iPhone® app

420    (Compass) appears to be superior to UG. Clinicians may find the iPhone® app (Compass) offers

421    greater convenience and efficiency than the UG, meaning that it could be introduced into practice

422    with confidence that it provides reliable measurements both within and between raters.

423    Considering the levels of agreement are clinically unacceptable the devices should not be used

424    interchangeably for initial and follow up measurements.

425 **REFERENCES**

426 Blanch P. 2004. Conservative management of shoulder pain in swimming. *Physical Therapy in*
427     *Sport* 5:109-124. 10.1016/j.ptsp.2004.05.002
428 Bland JM, and Altman DG. 2010. Statistical methods for assessing agreement between two
429     methods of clinical measurement. *International journal of nursing studies* 47:931-936.
430     https://doi.org/10.1016/j.ijnurstu.2009.10.001
431 Burdett RG, Brown KE, and Fall MP. 1986. Reliability and validity of four instruments for
432     measuring lumbar spine and pelvic positions. *Physical Therapy* 66:677-684.
433 Dixon-Warren S. 2012. Motion Sensing in the iPhone 4: Electronic Compass. *Available at*
434     http://www.memsjournal.com/2011/02/motion-sensing-in-the-iphone-4-electronic-
435     compass.html (accessed 15th Dec 2016).
436 Drumm J, and Swiegers M. 2015. Mobile Consumer Survey 2015 – The Australian Cut. *Available*
437     *at* http://landing.deloitte.com.au/rs/761-IBL-328/images/deloitte-au-tmt-mobile-
438     consumer-survey-2015-291015.pdf.
439 Furness J. 2015. Musculoskeletal and physiological profile of elite and recreational surfers :
440     injuries and sports specific screening. Thesis (PhD.) -- Bond University, 2015.
441 Furness J, Climstein M, Sheppard JM, Abbott A, and Hing W. 2016. Clinical methods to quantify
442     trunk mobility in an elite male surfing population. *Physical Therapy in Sport* 19:28-35.
443     http://dx.doi.org/10.1016/j.ptsp.2015.09.003
444 Furness J, Hing W, Abbott A, Walsh J, Sheppard JM, and Climstein M. 2014. Retrospective
445     analysis of chronic injuries in recreational and competitive surfers: injury location, type,
446     and mechanism. *International Journal of Aquatic Research and Education* 8:277-287.
447     10.1123/ijare.2013-0032
448 Gajdosik RL, and Bohannon RW. 1987. Clinical measurement of range of motion. Review of
449     goniometry emphasizing reliability and validity. *Physical Therapy* 67:1867-1872.
450 Iveson BD, McLaughlin SL, Todd RH, and Gerber JP. 2010. Reliability and exploration of the side-
451     lying thoraco-lumbar rotation measurement (STRM). North American Journal of Sports
452     Physical Therapy. p 201-207.
453 Jenny JY, Bureggah A, and Diesinger Y. 2016. Measurement of the knee flexion angle with
454     smartphone applications: Which technology is better? *Knee Surgery, Sports*
455     *Traumatology, Arthroscopy* 24:2874-2877. 10.1007/s00167-015-3537-4
456 Johnson KD, Kim K-M, Yu B-K, Saliba SA, and Grindstaff TL. 2012. Reliability of thoracic spine
457     rotation range of motion measurements in healthy adults. Journal of Athletic Training.
458     Dallas: National Athletic Trainers Association. p 52.
459 Jones A, Sealey R, Crowe M, and Gordon S. 2014. Concurrent validity and reliability of the Simple
460     Goniometer iPhone app compared with the Universal Goniometer. *Physiother Theory*
461     *Pract* 30:512-516. 10.3109/09593985.2014.900835
462 Kolber MJ, Fuller C, Marshall J, Wright A, and Hanney WJ. 2012. The reliability and concurrent
463     validity of scapular plane shoulder elevation measurements using a digital inclinometer
464     and goniometer. *Physiotherapy Theory & Practice* 28:161-168.
465     10.3109/09593985.2011.574203

466    Lam GC, Hill DL, Le LH, Raso JV, and Lou EH. 2008. Vertebral rotation measurement: a summary
467        and comparison of common radiographic and CT methods. *Scoliosis* 3:16. 10.1186/1748-
468        7161-3-16
469    Lewis JS, and Valentine RE. 2010. Clinical measurement of the thoracic kyphosis. A study of the
470        intra-rater reliability in subjects with and without shoulder pain. *BMC Musculoskeletal*
471        *Disorders* 11:39-39. 10.1186/1471-2474-11-39
472    Lexell JE, and Downham DY. 2005. How to assess the reliability of measurements in
473        rehabilitation. *American Journal of Physical Medicine & Rehabilitation* 84:719-723.
474        10.1097/01.phm.0000176452.17771.20
475    Mourcou Q, Fleury A, Diot B, Franco C, #xe9, line, and Vuillerme N. 2015. Mobile Phone-Based
476        Joint Angle Measurement for Functional Assessment and Rehabilitation of
477        Proprioception. *BioMed Research International* 2015:15. 10.1155/2015/328142
478    Nadalen MD. 2016. Basic Injury Prevention Concepts. *Available at* http://www.acsm.org/public-
479        information/articles/2016/10/07/basic-injury-prevention-concepts.
480    Nitschke J, Nattrass C, Disler P, PhD F, Chou M, and Ooi K. 1999. Reliability of the American
481        Medical Association Guides' Model for Measuring Spinal Range of Motion: Its Implication
482        for Whole-Person Impairment Rating. *Spine* 24:262-268.
483    Norkin C, and White J. 2016. *Measurement Of Joint Motion: A Guide To Goniometry*.
484        Philadelphia: F. A. Davis Company.
485    O'Gorman H, and Jull G. 1987. Thoracic kyphosis and mobility: the effect of age. Physiotherapy
486        Practice. p 154-162.
487    Pollard H, and Fernandez M. 2004. Spinal Musculoskeletal Injuries Associated with Swimming: A
488        Discussion of Technique. *Australasian Chiropractic & Osteopathy* 12:72-80.
489    Salamh PA, and Kolber M. 2014. The reliability, minimal detectable change and concurrent
490        validity of a gravity-based bubble inclinometer and iphone application for measuring
491        standing lumbar lordosis. *Physiotherapy Theory & Practice* 30:62-67.
492        10.3109/09593985.2013.800174
493    Schram B. 2015. Stand up paddle boarding : an analysis of a new sport and recreational activity.
494        Thesis (PhD.) -- Bond University, 2015.
495    Shrout PE, and Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability.
496        *Psychological Bulletin* 86:420-428. 10.1037/0033-2909.86.2.420
497    Spurrier D. 2006. Pre-season screening and injury prevention. *Sports Coach* 29.
498    Suresh KP. 2011. An overview of randomization techniques: An unbiased assessment of outcome
499        in clinical research. *Journal of Human Reproductive Sciences* 4:8-11. 10.4103/0974-
500        1208.82352
501    Tousignant-Laflamme Y, Boutin N, Dion AM, and Vallée C-A. 2013. Reliability and criterion validity
502        of two applications of the iPhone™ to measure cervical range of motion in healthy
503        participants. *J Neuroeng Rehabil* 10:69. 10.1186/1743-0003-10-69
504    Trevethan R. 2017. Intraclass correlation coeffients: clearing the air, extending some cautions,
505        and making some requests. *Health Services and Outcomes Research Methodology*
506        17:127-143. 10.1007/s10742-016-0156-6
507    Wilson TV, and Fenlon W. 2007. How the iPhone works. *Available at*
508        http://electronics.howstuffworks.com/iphone.htm (accessed April 24 2017).
509    Youdas JW, Carey JR, and Garrett TR. 1991. Reliability of measurements of cervical spine range of
510        motion--comparison of three methods. *Physical Therapy* 71:98-104; .

511   **APPENDICES**

512   *Appendix A*

513   1. "Keep your feet flat on the floor, hold the ball in between your legs firm enough that it

514      won't drop."

515   2. "Cross your arms over your chest, hold the bar against your collar bones with your

516      fingertips and maintain the pressure of the bar on your chest as you perform the

517      movement."

518   3. "Rotate to the left/right as far as you comfortably can while keeping your head in line

519      with your shoulders."

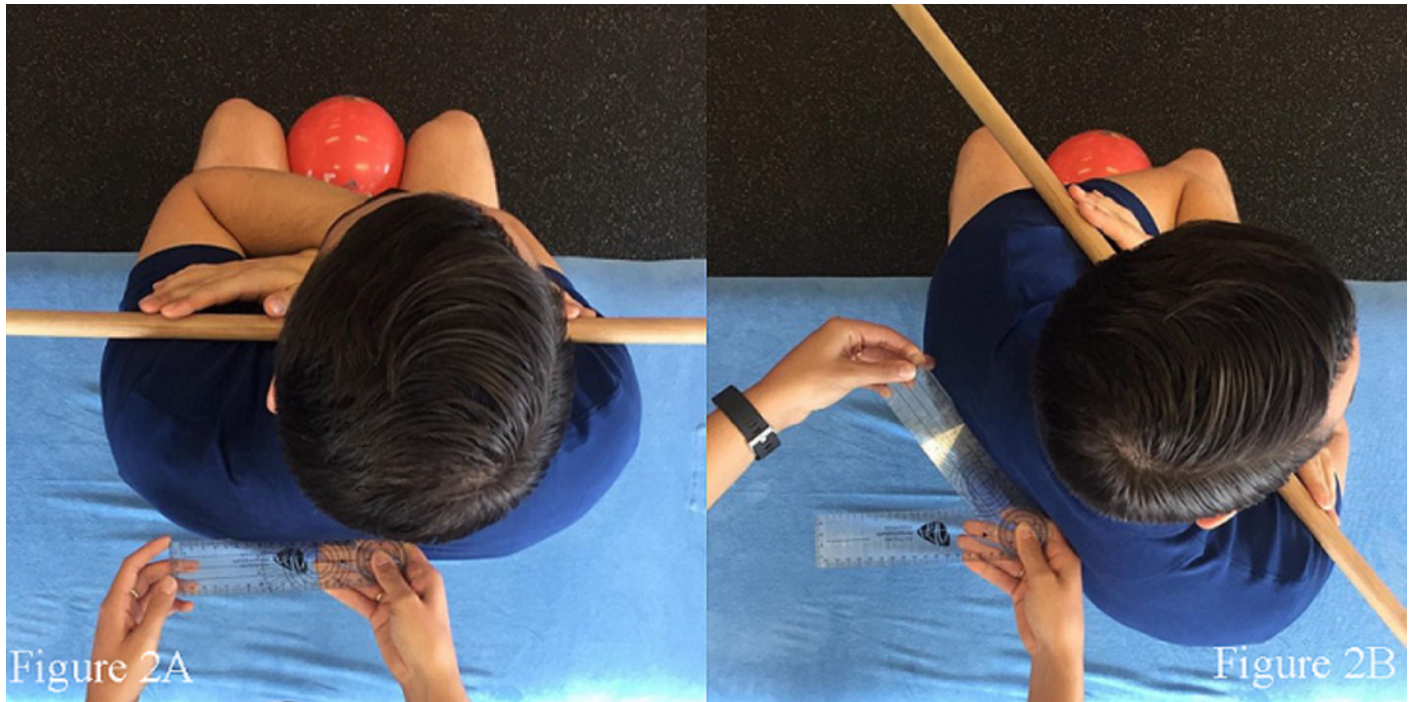
# Figure 1

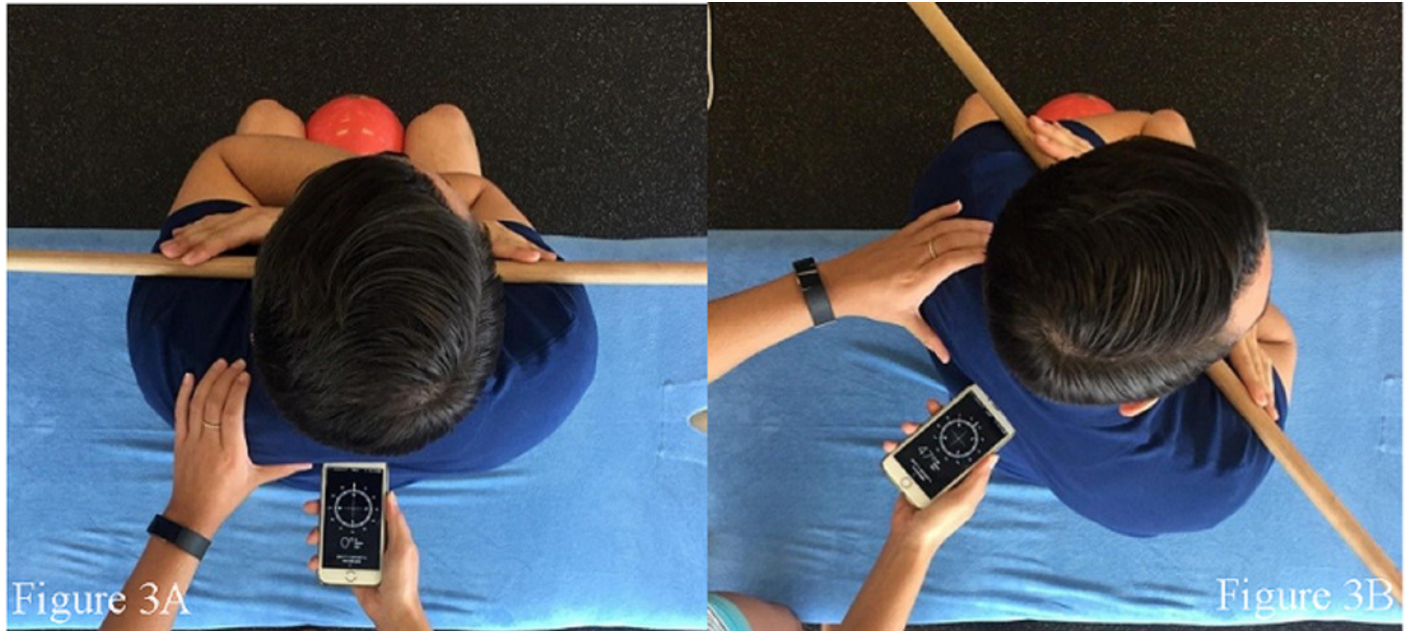
Starting position for seated rotation bar-in-front

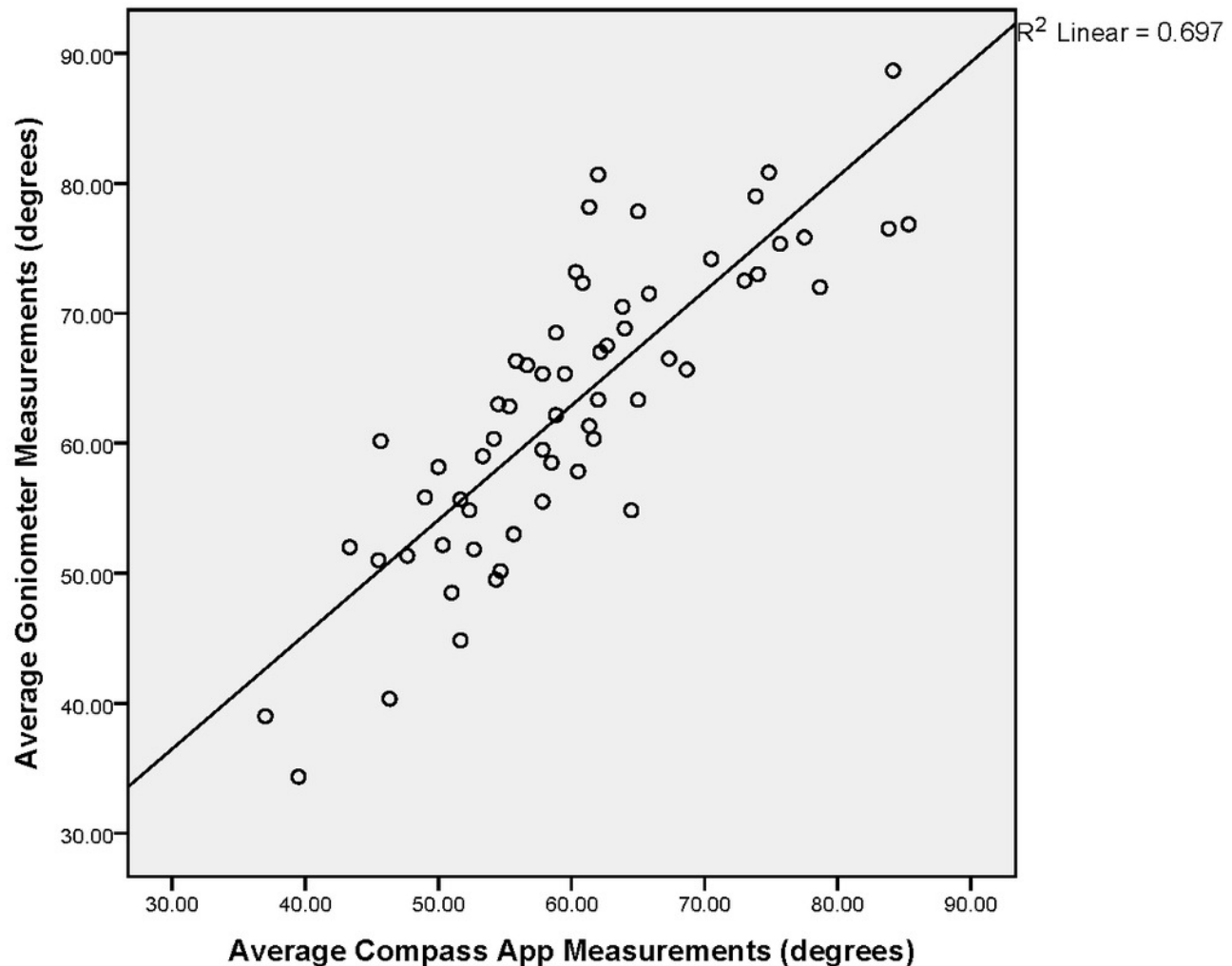
Figure 1A

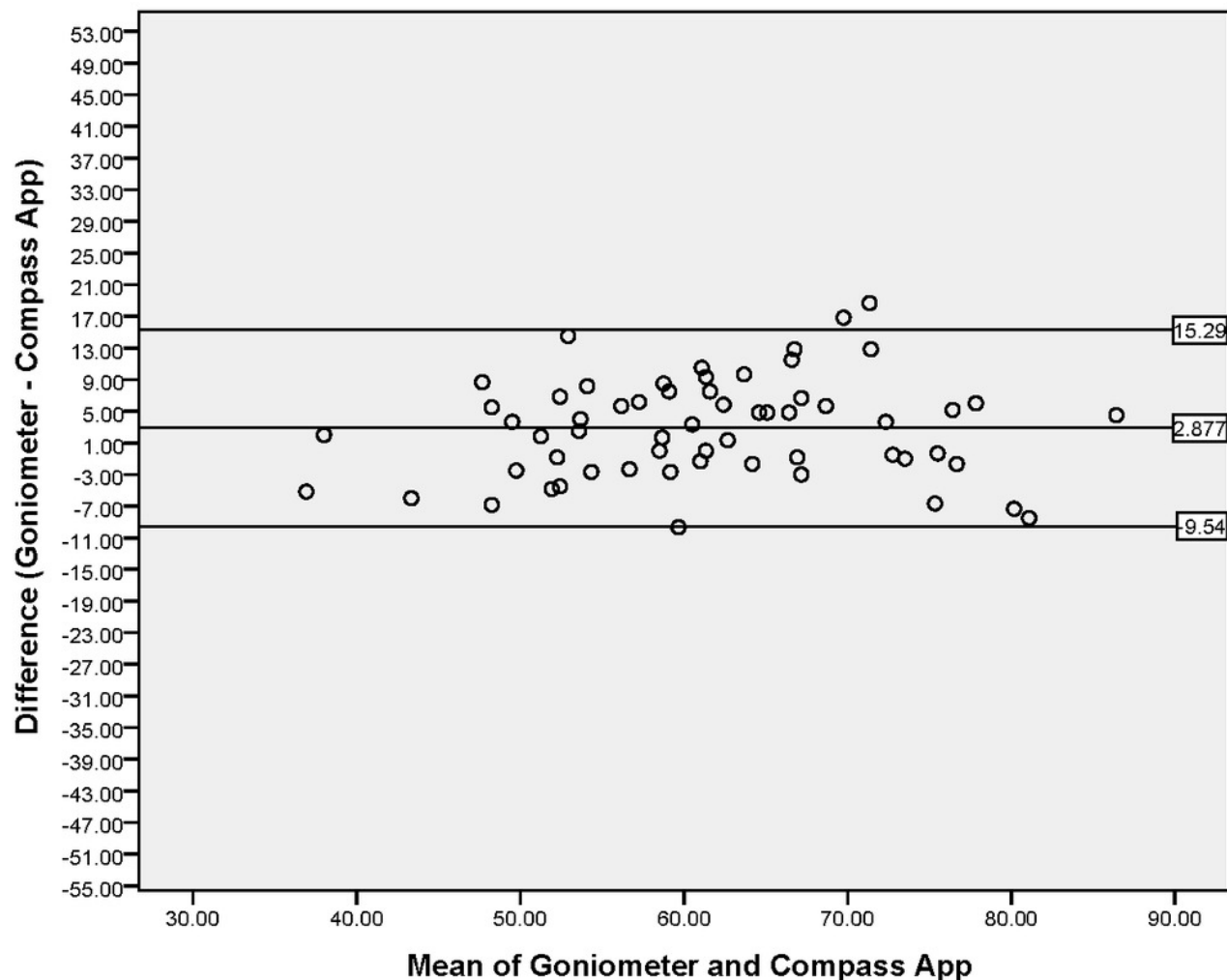
Figure 1B

# Figure 2

Measurement of thoracic rotation with the universal goniometer

# Figure 3

Measurement of thoracic rotation with the iPhone Compass App

# Figure 4

Scatterplot Showing Relationship between iPhone® Compass app and Goniometer for thoracic rotation ROM

# Figure 5

Bland–Altman Plot indicating mean difference and 95% limits of agreement between measurements from the iPhone® Compass app and Goniometer for thoracic rotation (°)

Descriptive Data for Thoracic Rotation

| | n =30 | $\bar{x}$ (°) | SD |
|---|---|---|---|
| **Left** | Goniometer | 63.5 | 11.4 |
| | Compass App | 60.3 | 11.7 |
| **Right** | Goniometer | 62.5 | 12.2 |
| | Compass App | 59.9 | 11.6 |
| **Average** | Goniometer | 63.0 | 11.3 |
| | Compass App | 60.1 | 10.7 |

$\bar{x}$ = mean value, SD = standard deviation, n = number of subjects

Intra-rater reliability for examiner 1 and 2 comparing the UG and the iPhone® Compass application

| Rater #1 (three measures each side) | | | | | | |
|---|---|---|---|---|---|---|
| | Right | | | Left | | |
| | ICC (95% CI) | SEM (º) | SRD (º) | ICC (95% CI) | SEM (º) | SRD (º) |
| **Goniometer** | 0.97 (0.95, 0.98) | 2.95 | 11.56 | 0.98 (0.97, 0.99) | 2.23 | 8.74 |
| **Compass** | 0.97 (0.96, 0.99) | 2.90 | 11.36 | 0.98 (0.97, 0.99) | 2.44 | 9.56 |
| Rater #2 (three measure each side) | | | | | | |
| | Right | | | Left | | |
| | ICC (95% CI) | SEM (º) | SRD (º) | ICC (95% CI) | SEM (º) | SRD (º) |
| **Goniometer** | 0.97 (0.94, 0.98) | 3.87 | 15.17 | 0.94 (0.88, 0.97) | 4.36 | 17.09 |
| **Compass** | 0.96 (0.93, 0.98) | 3.79 | 14.73 | 0.97 (0.95, 0.98) | 2.98 | 11.68 |

*ICC = Intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SRD = standard real difference.*

Inter-rater reliability for the UG and the iPhone® Compass application

| | Right (average of three measures) | | | Left (average of three measures) | | |
|---|---|---|---|---|---|---|
| | ICC (95% CI) | SEM (°) | SRD (°) | ICC (95% CI) | SEM (°) | SRD (°) |
| Goniometer | 0.85 (0.57, 0.94) | 6.33 | 17.46 | 0.72 (0.21, 0.88) | 7.85 | 21.76 |
| Compass | 0.87 (0.74, 0.91) | 5.46 | 15.13 | 0.89 (0.77, 0.95) | 5.17 | 14.33 |

*ICC = Intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SRD = standard real difference*