Rater reliability and construct validity of a mobile application for posture analysis

Kimberly A. Szucs, PhD, OTR/L1*, Elena V. Donoso Brown, PhD, OTR/L1

1) Department of Occupational Therapy, Duquesne University: 600 Forbes Avenue, Pittsburgh, PA, USA

Abstract. [Purpose] Measurement of posture is important for those with a clinical diagnosis as well as researchers aiming to understand the impact of faulty postures on the development of musculoskeletal disorders. A reliable, cost-effective and low tech posture measure may be beneficial for research and clinical applications. The purpose of this study was to determine rater reliability and construct validity of a posture screening mobile application in healthy young adults. [Subjects and Methods] Pictures of subjects were taken in three standing positions. Two raters independently digitized the static standing posture image twice. The app calculated posture variables, including sagittal and coronal plane translations and angulations. Intra- and inter-rater reliability were calculated using the appropriate ICC models for complete agreement. Construct validity was determined through comparison of known groups using repeated measures ANOVA. [Results] Intra-rater reliability ranged from 0.71 to 0.99. Inter-rater reliability was good to excellent for all translations. ICCs were stronger for translations versus angulations. The construct validity analysis found that the app was able to detect the change in the four variables selected. [Conclusion] The posture mobile application has demonstrated strong rater reliability and preliminary evidence of construct validity. This application may have utility in clinical and research settings.

Key words: Reliability, Construct validity, Posture

INTRODUCTION

Posture is defined in general as the position of one’s body in either a sitting or a standing position. The measurement of posture is common in many health professions as atypical postures are often present with certain diagnoses. For example, in children and adolescents there is a risk for scoliosis or curved spine1 while persons with Parkinson’s disease3 are noted to have a posture of flexion at the head and trunk. Repetitive or prolonged positioning in an atypical posture may also present as a risk factor for the development of musculoskeletal disorders, including trapezius myalgia and subacromial impingement syndrome3, 4. Therefore, measurement of posture is critical for those with a clinical diagnosis as well as researchers aiming to understand the impact of prolonged faulty postures on the development of musculoskeletal problems.

Current methods for the measurement of posture span from goniometry5 to photographic assessment with measurement of angles by hand or via computer software6, 7, to complex and expensive camera systems that can provide automated output, such as Bionix (Bionix, Toledo, OH, USA) or TrakStar (Ascension Technology Corp., Shelburne, VT, USA). Clinically, the goniometer is frequently used with good reliability to assess joint angles5, 8. This measurement method is inexpensive and simple but provides limited information regarding joint segment translations. More advanced measurement methods are potential options for clinicians, however, those that employ marker systems which are able to give highly precise readings are often too expensive, are time consuming to set up, and lack portability for the average clinic. In research settings, camera and marker systems are typically used for the assessment of posture secondary to their reliability and validity in capturing joint
angles and translations\textsuperscript{9–11}, however similar challenges such as expense are faced by junior faculty setting up their lab or researchers at smaller institutions.

Given the increase in available motion capture technology over the past two decades, more options that are cost effective and low tech are becoming accessible for clinical use\textsuperscript{12, 13}. For instance, there are now photographic posture assessment mobile applications that aim to provide clinicians and researchers with tools for measuring posture and documenting changes in posture over time. While mobile applications have the potential to increase the feasibility and ease of postural measurement for clinicians and researchers with a more streamlined and cost effective process, it is critical to ensure that the tool is reliable and valid.

Previous research on the PostureScreen Mobile\textsuperscript{®} application have found less than desirable inter- and test-retest reliability\textsuperscript{14, 15} and notable differences in measurement when compared to a gold standard 3-D system\textsuperscript{16}. Boland and colleagues\textsuperscript{14} recently reported ICC’s for inter-rater reliability ranging from 0.26 to 0.93 for eleven postural variables, including translations and angulations, using the PostureScreen Mobile\textsuperscript{®} app, while Kan et al.\textsuperscript{15} found ICC’s for inter-rater reliability between −0.12 and 0.96 across two data collection sessions. To improve rater reliability, Boland and colleagues recommended assessing clients who are minimally dressed. They also discussed difficulties in identifying certain landmarks, especially for examiners who may be inexperienced. The purpose of this study is to improve reliability when using this app by addressing the issues identified by Boland et al.\textsuperscript{14} to see if rater reliability can in fact be improved and if additional, not previously examined, posture variables can be digitized reliably. Furthermore, the construct validity through a comparison of known changes in posture will be evaluated. This analysis has not previously been conducted and would provide novel and useful information regarding the potential use of this app in detecting postural changes.

**SUBJECTS AND METHODS**

A sample of twenty healthy volunteers (17 females, 3 males) completed this study (21.1 ± 1.5 years, all right hand dominant). Exclusion criteria included history of or current shoulder, neck or back injury/pathology, acute musculoskeletal pain, and body mass index ≥25. All potential participants underwent a physical screening to confirm eligibility, which included an assessment of upper extremity (UE) and neck range of motion, UE strength, and tests for underlying neck and shoulder pathology. This study was approved by the University’s Institutional Review Board (Protocol No. 2015-08-2), and all subjects provided informed consent prior to starting data collection.

The PostureScreen Mobile\textsuperscript{®} app was used to assess baseline posture during standing. This app was specifically designed for health professionals who are interested in objectively evaluating posture, movement and body composition of patients (PostureCo; http://postureanalysis.com/mobile/). The app calculates posture variables using digitized anatomical landmarks from either 2 or 4 pictures of the individual, depending on the number of variables of interest. The device camera is used within the app to take pictures of subjects from various directions—anterior and posterior (coronal plane), left and right (sagittal plane). When taking the picture, the app shows a target-like display that turns green when the tablet is level. This helps ensure each image is taken at a level and consistent angle. After the picture is taken, the user is directed to crop each picture at the subject’s head and feet, giving a full view of the body. This ensures that the proportions of the subject are consistent between images.

Once each picture is captured, the app takes the user through the process of digitizing specified anatomical landmarks in order to produce anterior and lateral translations and angular displacements. During digitization, a grid is laid over the picture and an image of the anatomical landmark is inset into the picture to assist with landmark location. Following digitization, the app calculates the translations and angulations. Joint segment translations are calculated relative to the inferior segment (Fig. 1, Table 1). For example, translation of the head in the coronal plane (termed Head Anterior Translation in the app) is calculated by determining the offset between the midpoint of the eyes and the episternal notch (PostureCo; http://postureanalysis.com/mobile/). Joint angulations in the coronal plane are calculated based on the horizontal plane and joint...
To begin the postural assessment, subjects stood on a marked spot and they were instructed to stand in a typical, comfortable posture. Using the iPad camera, a research assistant took four pictures of each subject, from the anterior aspect (coronal plane), left/right aspects (sagittal plane), and posterior aspect (coronal plane). The research assistant stood on marked spots 8’ away from the subject for each picture. The height of the camera was not standardized as it was adjusted based on the subject’s height. However, guides provided within the app ensured that the entire body was included in the pictures being taken. Once the pictures of static standing were taken, subjects then were instructed to use their cell phone to complete two tasks: 1) texting a standardized message and 2) making a phone call. The phone call was made to a research assistant who engaged with a research assistant in the phone call and texting task during this entire duration. All postural assessments were engaged with a research assistant in the phone call and texting task during this entire duration. All postural assessments (both for static standing and phone tasks) were completed in one session.

Two raters were responsible for digitizing the same sets of pictures. This method was selected in order to focus on the placement of digital markers through the application and not include additional sources of error such as the taking and cropping of the photos specifically, which would be considered when evaluating test-retest reliability. To determine intra-rater reliability, each rater digitized the same set of photos from each participant twice. The time between the first and second digitization was set at two weeks. Prior to digitization, the raters discussed and identified the specific locations that would be digitized for each anatomical landmark indicated by the app. They did not discuss digitization of specific pictures or subjects. The output from the app was exported and values were compiled into a spreadsheet.

For the testing of rater reliability, out of the thirty-nine possible variables, the authors focused on twenty variables that included both translations and angulations from all four views. Intra-rater reliability was calculated using an ICC (3,1) mixed model for complete agreement, while inter-rater reliability was calculated using an ICC (2,1) random model for complete agreement\(^1\). Inter-rater reliability was calculated after intra-rater reliability was established using the first trial of both raters (KS and EDB). According to Portney and Watkins\(^1\), results within the range 0.00–0.25 represent no to little relationship, 0.25–0.50 represent a fair relationship, 0.50–0.75 represent a moderate relationship, and values >0.75 are considered good to excellent.

In order to analyze the data for construct validity, data from a single rater (KS) was used from three separate positions: 1) static standing, 2) device use while texting, 3) device use while making a phone call. A total of seven variables (Head Anterior and Right Lateral Translation, Head Anterior and Right Lateral Angulation, Shoulder Anterior and Right Lateral

### Table 1. Calculation of posture variables

| Plane of view | Variables of interest | Points needed for variable calculation |
|---------------|-----------------------|---------------------------------------|
| Anterior coronal | Anterior head translation | Bilateral pupils, sternal notch |
| | Shoulder anterior translation | Bilateral acromioclavicular joints, bilateral T8 ribs |
| | Rib anterior translation | Bilateral T8 ribs, bilateral anterior superior iliac spines |
| | Hip anterior translation | Bilateral anterior superior iliac spines, midpoint of bilateral anterior ankles |
| | Head anterior angulation | Bilateral pupils, Sternal notch |
| | Shoulder anterior angulation | Bilateral acromioclavicular joints |
| | Hip anterior angulation | Bilateral anterior superior iliac spines |
| Sagittal | Head right/Left lateral translation | External auditory meatus, Acromioclavicular joint |
| | Shoulder right/Left lateral translation | Acromioclavicular joint, Greater trochanter of femur |
| | Hip right lateral translation | Greater trochanter of femur, Lateral epicondyle of femur |
| | Head right/Left lateral angulation | External auditory meatus, Acromioclavicular joint |
| Posterior coronal | Head posterior translation | Bilateral inferior ear lobe, bilateral acromioclavicular joint |
| | Shoulder posterior translation | Bilateral acromioclavicular joint, bilateral T8 ribs |
| | Rib posterior translation | Bilateral T8 ribs, bilateral posterior superior iliac spines |
| | Hip posterior translation | Bilateral posterior superior iliac spines, bilateral Achilles tendon |
| | Shoulder posterior angulation | Bilateral acromioclavicular joint |

Joint translations are calculated relative to the inferior segment and angulations are calculated with two points and either the horizontal or vertical plane depending on the plane of view.
Translation, and Rib Anterior Translation) were selected and a priori hypotheses were set for whether or not change in that variable was expected given the change in task. Changes were anticipated in head angulation and translation variables but not in shoulder and rib translations. Data were first checked for outliers using the box plot method evaluating for values greater than 1.5 box-lengths from the edge of the box. A total of 5 outliers were identified. All outliers were checked for measure-
ment accuracy and two were found to have been data entry errors. The other three outliers were removed from the analysis in order to provide a more conservative estimate of differences between groups. Values for each variable in each position were normally distributed, as assessed by Shapiro-Wilk’s test (p>0.05). Then a one way repeated measures ANOVA with a Greenhouse Geisser adjustment for sphericity was run for each variable with post-hoc simple contrasts using Bonferroni adjustments. Alpha level was set at ≤0.025 for statistical significance.

**RESULTS**

All subjects included in the study were healthy and free of pathology in their shoulders or necks. Average height and standard deviation for subjects was 65.7 ± 2.2 inches and average weight and standard deviation was 137.6 ± 17.2 pounds. Calculated average BMI of all subjects was 22.2 ± 2.2.

Intra-rater reliability for rater 1 (KS) had ICC values from 0.76 to 0.99 with an average of 0.93. Rater 2 (EDB) showed similar correlations with ICC values ranging from 0.71 to 0.99 with an average ICC of 0.92. ICCs appeared to be stronger for translations versus angulations for both raters (Tables 2 & 3). Inter-rater reliability was 0.85 to 0.99 for all translations with an average ICC of 0.95 (Table 2). Inter-rater reliability was less strong ranging from 0.66 to 0.98 for angulation variables. Head Posterior Angulation demonstrated the weakest relationship having an ICC of 0.66. However, the average ICC for angulations was 0.97 (Table 3).

All four variables that measured head position demonstrated statistically significant differences across the three tasks (Table 4). Post hoc contrasts showed statistically significant differences between standing and texting and standing and cell phone call for three out of four variables, with the other two trending toward significance. Two out of the three variables where no change was expected did not demonstrate a statistically significant change, with the measurement of Shoulder Right Lateral Translation demonstrating a statistically significant difference across all three positions and in post hoc contrasts (Table 4).

| Table 2. Intra-rater and inter-rater reliability translation variables |
|---------------------------------------------------------------|
| **Variables (inches)** | **Intra-rater KS** | **Intra-rater EDB** | **Interrater** |
|------------------------|------------------|------------------|-------------|
|                       | Mean (SD) 1      | Mean (SD) 2      | ICC         |
| Head anterior translation | 0.30 (0.318)     | 0.27 (0.312)     | 0.93        |
| Head right lateral translation | 1.79 (1.04)     | 1.80 (1.03)     | 0.99        |
| Head posterior translation | 0.09 (0.38)     | 0.12 (0.31)     | 0.91        |
| Head left lateral translation | 1.91 (1.07)     | 1.87 (1.06)     | 0.99        |
| Shoulder anterior translation | 0.09 (0.37)     | 0.06 (0.40)     | 0.92        |
| Shoulder right posterior translation | -0.20 (1.55) | -0.39 (1.64) | 0.94        |
| Shoulder posterior translation | -0.10 (0.38)    | -0.09 (0.33)    | 0.92        |
| Shoulder left lateral translation | -0.42 (1.34)    | -0.52 (1.42)    | 0.98        |
| Rib anterior translation | -0.20 (0.49)    | -0.18 (0.49)    | 0.92        |
| Rib posterior translation | 0.01 (0.61)     | 0.02 (0.65)     | 0.97        |
| Hip anterior translation | -0.01 (0.57)    | -0.005 (0.58)   | 0.99        |
| Hip right posterior translation | 0.33 (1.35)     | 0.56 (1.36)     | 0.92        |
| Hip posterior translation | 0.09 (0.75)     | 0.03 (0.75)     | 0.98        |

| Table 3. Intra-rater and inter-rater reliability angulation variables |
|---------------------------------------------------------------|
| **Variable (angles)** | **Intra-rater KS** | **Intra-rater EDB** | **Interrater** |
|------------------------|------------------|------------------|-------------|
|                       | Mean (SD) 1      | Mean (SD) 2      | ICC         |
| Head anterior angulation | 0.64 (3.59)     | 1.70 (3.15)     | 0.82        |
| Head right lateral angulation | 13.75 (7.87)   | 13.84 (7.57)    | 0.99        |
| Head posterior angulation | 1.04 (3.10)     | 0.41 (3.17)     | 0.90        |
| Head left lateral angulation | 13.55 (8.62)   | 14.70 (8.44)    | 0.76        |
| Shoulder anterior angulation | 1.20 (1.71)     | 1.35 (1.64)     | 0.95        |
| Shoulder posterior angulation | 1.09 (2.32)     | 0.89 (2.06)     | 0.95        |
| Hip anterior angulation | -0.77 (2.36)    | -1.01 (2.08)    | 0.77        |
The purpose of this study was to determine the rater reliability and construct validity of an app that was developed for the clinical assessment of posture. While previous studies have explored the reliability of a few posture variables using this app, the findings of this study are novel in that additional posture variables of interest were examined as was the construct validity of variables. Additionally, we were able to use recommendations from previous research to improve the procedures for using this app. Based on the results of this study, the PostureScreen Mobile® app, when using defined procedures and markers, has been found to be a reliable method for assessing posture within and across raters. For intra-rater reliability, the majority of posture variables showed good to excellent reliability (>0.75). The few variables that were within the moderate to good reliability range (0.50–0.75) were angulation variables. A similar pattern in reliability was observed for inter-rater measures, with Head Posterior Angulation and Head Left Lateral Angulation demonstrating moderate to good agreement. These reliability values are consistent with other widely used posture measures, including goniometers and inclinometers, both common clinical tools, and more advanced measures, such as electromagnetic tracking systems.

Based on the results from the known group analysis, this mobile app is able to capture changes in posture variables when they exist. For the four variables where change across tasks was expected, there was a statistically significant difference, with the exception of Head Anterior Translation when making a phone call and Head Anterior Translation when texting, though this trended towards significance at p=0.06 and p=0.05 respectively. One unexpected finding was the statistically significant difference across the three posture tasks for Shoulder Right Lateral Translation. Prior to data analysis, the authors did not anticipate much difference in the shoulder position from static standing while subjects used their phones. However, upon further inspection of the pictures taken during these tasks, it is evident that the shoulder is positioned posterior to the hip while subjects were both texting and making a phone call. While this finding did not support the a priori hypothesis on this specific posture variable, it does support the overall hypothesis that this app is able to detect change where it exists. Additionally, the other two variables where no changes were expected (Shoulder Anterior Translation and Rib Anterior Translation) did not show a statistically significant difference between posture tasks, also supporting this hypothesis.

Biomechanics research frequently utilizes electromagnetic tracking sensors or optic sensors with multiple camera systems for data collection methods, however these systems are expensive and may not be a feasible option at smaller academic institutions. This app presents a cost-effective alternative for both clinics and research laboratories for measuring and describing baseline posture. The initial cost to purchase the app is nominal in comparison to traditional biomechanics instrumentation. While the app costs more than traditional clinical measures of posture, such as goniometers and inclinometers, it offers an analysis in multiple views and increased quantity of variables with additional posture variables not available with clinical tools (e.g., joint segment translations).

Beyond the affordable cost, another benefit of using this app is its relative ease of use. The raters in this study found this mobile application to be a user-friendly method to assess posture. The company’s website provides background and instructional information regarding the app’s uses. Once installed and opened, the user interface is intuitive and the process for taking the client pictures is easy to follow. In the recent study by Boland et al. [14], the postural assessment was completed by two inexperienced undergraduate students with no reports of difficulty. All posture variables are calculated by the app and provided in table format. While instructions are provided at each step of the posture screen process, there are instances where these instructions are not clear. During this study, the researchers found a lack of detailed instructions regarding where exactly on the anatomical landmark to digitize. Only basic instructions were provided within the app, such as ‘Place marker on the Right Shoulder’ with an inset picture showing the marker on an area near or at the acromioclavicular joint. To increase agreement with digitizing, the raters discussed exact locations for digitization prior to beginning this process. In addition, reflective markers were placed on the landmarks to assist with locating them on the pictures, which also potentially contributed to greater consistency. Many of the landmarks are difficult to directly identify from the pictures (i.e., greater trochanter of the femur). Though not specifically recommended by the app’s manufacturers, the authors suggest that in order to be consistent
and accurate in the digitization process, it is necessary to identify specific landmarks with reflective markers prior to taking the picture. This may present as a time constraint for already busy clinicians, though not all anatomical landmarks, such as the ear tragus, would require a marker. As Boland and colleagues noted, some landmarks are more difficult for inexperienced examiners to identify during digitization therefore this recommendation may improve reliability across all levels of examiners by clearly marking these landmarks in advance of digitization.

This study demonstrated good to excellent intra- and inter-rater reliability when digitizing images using a mobile posture application for both segment translations and angulations. The app also demonstrated construct validity across known positions. This app provides greater number and type of posture related variables than goniometric or inclinometer measurements and is more cost-effective and user friendly than advanced marker and camera systems. The outcomes of this study provide support for the use of markers and minimal clothing when measuring baseline posture in clinical and research settings. The issue of measurement reliability is an important and necessary first step in determining a tool’s utility, however, further research into other psychometric properties, like test-retest reliability, which would assess the entire assessment process over two time points and concurrent validity with advanced camera systems, is warranted.

In this study, all data were collected from healthy, young subjects. We recognize this as a possible limitation as the results may not be generalizable to a clinical population. In addition, efforts were made to standardize all parts of the data collection procedures, the height from which pictures were taken was not standardized but instead adjusted based on the height of the individual subject. The height of the camera could be standardized to further reduce potential sources of error. Furthermore, the procedures in this study do not fully replicate what would occur in clinical practice or cover all sources of error, however these findings do support that procedures can be used to improve rater reliability when this is applied in the clinic.

Conflicts of interest
The authors declare no conflicts of interest.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

1) Nichelle da Rosa B, Noll M, Sedrez JA, et al.: Monitoring the prevalence of postural changes in schoolchildren. J Phys Ther Sci, 2016, 28: 326–331. [Medline] [CrossRef]
2) Bloem BR, Marinus J, Almeida Q, et al. Movement Disorders Society Rating Scales Committee: Measurement instruments to assess posture, gait, and balance in Parkinson’s disease: Critique and recommendations. Mov Disord, 2016, 31: 1342–1355. [Medline] [CrossRef]
3) Shin SJ, Yoo WG: Changes in cervical range of motion, flexion-relaxation ratio and pain with visual display terminal work. Work, 2014, 47: 261–265. [Medline]
4) Lewis JS, Green A, Wright C: Subacromial impingement syndrome: the role of posture and muscle imbalance. J Shoulder Elbow Surg, 2005, 14: 385–392. [Medline] [CrossRef]
5) Youdas JW, Carey JR, Garrett TR: Reliability of measurements of cervical spine range of motion—comparison of three methods. Phys Ther, 1991, 71: 98–104, discussion 105–106. [Medline] [CrossRef]
6) do Rosário JL: Photographic analysis of human posture: a literature review. J Bodyw Mov Ther, 2014, 18: 56–61. [Medline] [CrossRef]
7) Hazar Z, Karabicak GO, Tiftikci U: Reliability of photographic posture analysis of adolescents. J Phys Ther Sci, 2015, 27: 3123–3126. [Medline] [CrossRef]
8) Wolfenberger VA, Bui Q, Batenchuk GB: A comparison of methods of evaluating cervical range of motion. J Manipulative Physiol Ther, 2002, 25: 154–160. [Medline] [CrossRef]
9) Koerhuis CL, Winters JC, van der Helm FC, et al.: Neck mobility measurement by means of the ‘Flock of Birds’ electromagnetic tracking system. Clin Biomech (Bristol, Avon), 2003, 18: 14–18. [Medline] [CrossRef]
10) Karduna AR, McClure PW, Michener LA, et al.: Dynamic measurements of three-dimensional scapular kinematics: a validation study. J Biomech Eng, 2001, 123: 184–190. [Medline] [CrossRef]
11) Teece RM, Lundin JB, Lloyd AS, et al.: Three-dimensional acromioclavicular joint motions during elevation of the arm. J Orthop Sports Phys Ther, 2008, 38: 181–190. [Medline] [CrossRef]
12) Cuesta-Vargas AI, Roldán-Jiménez C: Validity and reliability of arm abduction angle measured on smartphone: a cross-sectional study. BMC Musculoskelet Disord, 2016, 17: 93. [Medline] [CrossRef]
13) Vohralik SL, Bowen AR, Burns J, et al.: Reliability and validity of a smartphone app to measure joint range. Am J Phys Med Rehabil, 2015, 94: 325–330. [Medline] [CrossRef]
14) Boland DM, Neufeld EV, Ruddell J, et al.: Inter- and intra-rater agreement of static posture analysis using a mobile application. J Phys Ther Sci, 2016, 28: 3398–3402. [Medline] [CrossRef]
15) Kan S, Fortuna J, Nur S, et al.: Can a mobile application be used to assess postural alignment in clinic? Canadian Society for Biomechanics, 2016, Hamilton, ON.
16) Hopkins BC: Validity of PostureScreen Mobile® in the measurement of standing posture. Prove: Brigham Young University, 2014.
17) Portney LG, Watkins MP: Foundations of clinical research: application to practice, 2nd ed. Upper Saddle River: Pearson Prentice Hall, 2000.
18) Rousson V, Gasser T, Seifert B: Assessing intra-rater, inter-rater and test-retest reliability of continuous measurements. Stat Med, 2002, 21: 3431–3446. [Medline] [CrossRef]