Reliability of Visual Assessment of Forward Head Posture in Standing

Andrew Aitken

A research project submitted in partial requirement for the degree of Master of Osteopathy, Unitec New Zealand, 2008.
DECLARATION

Name of candidate: Andrew Aitken

This dissertation is submitted in partial fulfilment for the requirements for the Unitec degree of Master of Osteopathy

The regulations for the degree are set out in the MOst Programme Schedule and are elaborated in the course handbook.

Candidate’s declaration
I confirm that:

- This dissertation represents my own work;
- The contribution of any supervisors and others to the research and to the dissertation was consistent with the Unitec Code of Supervision.

Candidate: Date:

Supervisors’ declaration
I confirm that, to the best of my knowledge:

- The research was carried out and dissertation prepared under my direct supervision;
- Except where otherwise approved by the Board of Postgraduate Studies of Unitec, the research was conducted in accordance with the degree regulations and programme rules;
- The contribution made to the research by me, by other members of the supervisory team, by other members of staff of Unitec and by others was consistent with the Unitec code of supervision.

Supervisor: Date:
ABSTRACT

Background. The visual assessment of a patients’ standing posture is one of the first steps in the physical examination process as performed by osteopaths and other manual medicine practitioners. The assessment of forward head posture is an important part of this postural assessment. Only a small number of studies investigating the reliability of visual assessment of posture have been identified. This study may potentially add data that could be helpful in developing improved guidelines for assessing forward head posture.

This dissertation contains two main sections. Section one is a review of the literature that underlines the reliability of visual assessment of forward head posture. Section two consists of a manuscript that reports a study on the reliability of visual assessment of forward head posture in standing. The aim of the study was to investigate the intra and inter-rater reliability of visual assessment of forward head posture in standing as well as to determine the influence of osteopathic clinical experience on reliability of visual assessment of forward head posture.

Methods. A blinded test-retest design was used to examine the intra-rater and inter-rater reliability of the visual assessment of forward head posture in standing. Intra-rater reliability was investigated by having observers visually rate a video clip of the forward head posture of the same subject (n=60) twice. Inter-rater reliability was investigated by comparing results of 78 randomly selected observers comprised of 16 laypeople, 40 osteopathy students, and 22 osteopathic practitioners. The influence of clinical experience was investigated by comparing observers from these various groups.

Results. Intra-rater reliability across groups was only slightly better than would be expected by chance alone. Inter-rater reliability was universally statistically poor to fair, with only first and second year osteopathy students providing a moderate level of reliability.

Conclusions. The results call into question the statistical reliability of the visual assessment of forward head posture. More research is required in order to determine the variables that may confound statistical reliability of this commonly used clinical assessment technique.
ACKNOWLEDGMENTS

My sincere thanks are extended to the many people who assisted me throughout this research. In particular:

Robert Moran and Chris McGrath, for the time and effort spent supervising this research. Mirko Mirković for his help with developing and programming the interactive CD-ROM used in this study.

To all the patients from the Unitec New Zealand osteopathy clinic for participating in the study, and all the observers that gave up their valuable time.

Fionna Urlich for her help in organising the data collection room and assistance with recruiting subjects for the study.

To my family for all of their support.
# TABLE OF CONTENTS

DECLARATION .................................................................................................................. 2  
ABSTRACT ........................................................................................................................ 3  
ACKNOWLEDGMENTS ................................................................................................. 4  
TABLE OF CONTENTS ............................................................................................... 5  
LIST OF ABBREVIATIONS ......................................................................................... 6  
SECTION ONE: LITERATURE REVIEW .................................................................... 7  
  Introduction .................................................................................................................. 7  
  Literature Search ........................................................................................................ 8  
  Part One: Reliability ................................................................................................... 8  
  Part Two: Observation and Visual Assessment ........................................................ 9  
  Part Three: Skill Acquisition ....................................................................................... 10  
  Part Four: Posture ...................................................................................................... 11  
  Head and Neck Posture ............................................................................................. 11  
  Forward Head Posture ................................................................................................. 11  
  Rating of Forward Head Posture ................................................................................ 12  
  Part Five: Reliability of Visual Postural Assessment ................................................ 13  
    Reliability of Visual Postural Assessment after Stroke ........................................ 17  
    Reliability of Visual Assessment of Gait ............................................................... 17  
    The Influence of Experience on Reliability of Visual Assessment ................... 18  
    Validity of Visual Postural Assessment ................................................................. 19  
    Reliability of Postural Assessment Involving the Use of Digitising Systems ...... 20  
    Reliability of Postural Assessment Involving the Use of Measurement Devices ... 22  
  Part Six: Rationale for Current Study ..................................................................... 25  
REFERENCES ............................................................................................................... 27  
SECTION TWO: MANUSCRIPT ................................................................................. 31  
APPENDICES ............................................................................................................... 61
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>AROM</td>
<td>Active Range Of Motion</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>C7</td>
<td>Cervical vertebrae number seven</td>
</tr>
<tr>
<td>CROM</td>
<td>Cervical Range Of Motion</td>
</tr>
<tr>
<td>CV</td>
<td>Craniovertebral</td>
</tr>
<tr>
<td>FHP</td>
<td>Forward Head Posture</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-class Correlation Coefficient</td>
</tr>
<tr>
<td>ID</td>
<td>Identification</td>
</tr>
<tr>
<td>k</td>
<td>kappa coefficient</td>
</tr>
<tr>
<td>n</td>
<td>Number</td>
</tr>
<tr>
<td>PADS</td>
<td>Postural Analysis Digitising System</td>
</tr>
<tr>
<td>UG</td>
<td>Universal Goniometer</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
<tr>
<td>VE</td>
<td>Visual Estimation</td>
</tr>
</tbody>
</table>
LITERATURE REVIEW

Introduction
The purpose of this literature review is to identify previous research into the area of reliability of visual postural assessment, as well as provide the reader with background information on this topic.

In osteopathy, effective treatment is dependent upon the extent to which practitioners can perform reliable and accurate measures of a specific physical attribute (e.g. forward head posture) or function. Unreliable or inaccurate assessment confounds the use of diagnostic models, compromising the practitioner’s ability to make informed decisions regarding diagnosis, aetiology and treatment progression, and therefore complicates the effective prescription of treatment protocols (Peeler and Anderson, 2007).

Part one of this literature review begins by defining reliability. Part two discusses the process of observation and visual assessment, followed by part three that provides a description of the two main models of skill acquisition. Part four discusses posture in general, head and neck posture and forward head posture. Part five is a critical review of literature and reviews studies involving; reliability of visual postural assessment, reliability of visual postural assessment after stroke, reliability of visual postural assessment of gait, the effect of experience on reliability of visual postural assessment, validity of visual postural assessment, and the use of digitising and measurement devices for studying reliability of visual postural assessment. Part six presents the rationale for the study that is reported in Section 2 of the dissertation.
Literature Search
A thorough search of the National Library of Medicine (MEDLINE), Allied and Complimentary Medicine Database (AMED), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Osteopathic Literature Database (OSTMED), Ebsco Health Databases, Cochrane Databases, Physiotherapy Evidence Database (PEDro), and ScienceDirect electronic database was performed using combinations of the following key words and phrases; Reliability, Cervical Spine, Neck, Forward Head Posture, Posture, Visual Assessment, Observation, Experience, Inter-rater, Intra-rater. Only a small number of studies investigating the reliability of visual assessment were identified, thus indicating a need for more research into this area.

Part One: Reliability
This section will provide an overview of the importance of reliability in the clinical setting with particular emphasis on diagnosis. Reliability has been defined as:

“The extent to which a test measurement or device produces the same results with different investigators, observers, or administration of the test over time. If repeated use of the same measurement tool on the same sample produces the same consistent results, the measurement is considered reliable” Mosby (2002).

Typically in the manual medicine context, the process of diagnosis is informed from the information gathered through the medical interview and clinical examination, of which visual assessment is an important component. This informed process helps to reduce reliance upon expensive diagnostic and imaging equipment. This subsequently requires some assurance of the reliability of visual assessment, particularly if used to help form a diagnosis. Every ‘test’ used in clinical examination has certain sensitivity (i.e. the proportion of people with disease who have a positive test result) and specificity (i.e. the proportion of people without disease who have a negative test result). No test has 100% sensitivity and specificity. For example, in the health industry it may be sensible to risk exposing healthy patients to costly diagnostic imaging tests (low
specificity) in order to increase the chance of identifying nearly all unhealthy patients (high sensitivity). This potential for error increases uncertainty and is a reminder that all decisions evolve from considerations of probabilities. A numeric value of probability can be calculated, providing an indication of uncertainty (Seidel et al, 2003).

**Part Two: Observation and Visual Assessment**

This section reviews the process and science of observation and visual assessment.

Eye dominance appears to be a significant element in the accuracy of visual judgment and becoming aware of backgrounds; lighting is also important in eliminating optical illusions as a source of error. In order to make such judgments, Chaitow (2003) cites the work of Mitchell (1976) who lists the need to be able to 1) identify and discriminate colour hues and saturations; 2) quantify ‘rectilinear length measurements, angular measurements, curvilinear and arcuate shapes, and their radius of curvature’; 3) sense horizontal and vertical frames of reference in which to make quantitative judgments; 4) appreciate motion, absolute motion or subjective awareness of motion in relation to him/herself or relative motion of one thing to another thing; 5) Demonstrate depth perception and the ability to estimate length and proportion.

Observation is very often the first component of the physical examination process (Magee, 2002). Gross et al (2002) states; “the examination should begin in the waiting room before the patient is aware of being observed. Information regarding the degree of the patient’s pain, disability, level of functioning, posture, and gait can be observed”. More specifically, observation is looking for symmetry or asymmetry of a patients posture, their morphology, muscle balance and associated hypertrophy or atrophy, fasiculations or spasms, discolouration, swelling, masses, and pain behaviour etc. (Seidel et al, 2003). According to (Chaitow, 2003), visual assessment is very important for making effective and reliable assessments and clinical judgments. This study shall endeavour to add data regarding the reliability of visual assessment.
Part Three: Skill Acquisition
This section will provide a brief overview of the two main models of skill acquisition. Knowledge of the process of skill acquisition is pertinent to the current study, which involves observers with varying levels of experience in regards to visual assessment of forward head posture. Anecdotally, there is an expectation (among the general as well as medical population) that the greater amount of experience an observer has with visual assessment, the higher their level of skill acquisition, and therefore the more reliable they should be. Two models that support this assumption are that of Fitts and Posner (1967) and Dreyfus and Dreyfus (1986). Each shall be discussed.

Fitts and Posner (1967) identified three distinct stages of skill learning: 1) Cognitive stage (plan formation), 2) Associative stage (practice), and 3) Autonomous stage (automatic execution). Anderson (1982) further developed the three stages. The cognitive stage involves an initial encoding of the skill into a form sufficient to permit the learner to generate the desired behaviour to at least some crude approximation. The association stage involves the “smoothing out” of the skill performance. Errors in the initial understanding of the skill are gradually detected and eliminated. The autonomous stage is one of graded continued improvement in the performance of the skill. The improvements in this stage often continue indefinitely (Ericsson, 1996) cited in (Reznick and MacRae, 2006). Expert performance represents the highest level of skill acquisition and is the final result of a gradual improvement in performance through extended experience in a given domain (Reznick and MacRae, 2006). According to Ericsson (1996) most professionals reach a stable average level of performance and maintain this status for the rest of their careers. However, volume alone does not account for the skill level among practitioners, since variations in performance have been shown among practitioners with high and very high volumes of patients (Ericsson, 1996). Deliberate practice is a critical process for the development of mastery or expertise (Reznick and MacRae, 2006). Anderson (1982) claims that it requires at least 100 hours of learning and practice to acquire any significant cognitive skill to a reasonable degree of proficiency.
Dreyfus and Dreyfus’ (1986) five stage model of skill acquisition identifies levels from 1 to 5 and labels the levels as novice, advanced beginner, competent, proficient, and expert. Dreyfus and Dreyfus (1986) state that at a particular stage someone can always imitate the thought process characteristics of a higher stage but will perform badly when lacking practice and concrete experience. In relation to the current study, this would suggest that the more experienced osteopaths and those who utilise visual observation more readily would out perform those with less experience and less utilisation.

**Part Four: Posture**

This section provides definitions of posture, head and neck posture, plus defines and discusses the various methods used to measure forward head posture.

Posture may be defined as “the attitude of the body”(Simon et al, 1999). Ideal posture is when the body parts, muscles and bones, are aligned and work together in harmony, protecting the body from injury or progressive deformity, regardless of attitude (Simon et al, 1999). Griegel-Morris et al (1992) and Kendall et al (2005) state that proper posture is believed to be a state of musculoskeletal balance that involves a minimal amount of stress or strain on the body. Poor posture is basically an unhealthy but correctable habit resulting in misalignment of various body parts. These body parts are generally accepted to be at higher risk for injury or pain due to the increased strain that misalignment places on supporting structures.

*Head and Neck Posture*

Normal postural alignment of the head over the thorax in the sagittal plane has been described as the vertical alignment of the external auditory meatus over the acromioclavicular joint aligned with a vertical postural line (Seaman and Troyanovich, 2000; Yip et al, 2007). Anterior deviation from this position is considered poor posture.

*Forward Head Posture*

Forward head posture (FHP) means that the head is in an anterior position in relation to the postural line (Yip et al, 2007). According to Magee (2002), FHP
commonly affects the atlanto-occipital joint, the cervical spine, the temporomandibular joint, the scapulothoracic joint and the glenohumeral joint. Muscles commonly shortened are the levator scapulae, sternocleidomastoid, scalenes, suboccipital muscles, upper trapezius, pectoralis major and minor. Muscles commonly weakened are hyoid muscles, lower cervical and thoracic erector spinae, middle and lower trapezius and rhomboids. Braun and Amundson (1989) claim that FHP also may promote a decreased lordosis of the lower cervical spine resulting in nerve impingement and joint inflexibility. Additionally, the posterior cranial rotation of the head on the upper cervical spine that is associated with this FHP may be sufficient to compress the arteries and nerves exiting the skull suboccipitally. Local symptoms believed to be associated with FHP may include decreased range of neck motion, muscle stiffness or pain, and degenerative changes in the spine. Headaches, neck ache, and shoulder pain are common manifestations of these structural problems (Braun and Amundson, 1989). Chronic tension-type headaches, cervicogenic headache and tension-type headache have all been related to a smaller cranio-vertebral (CV) angle when compared to controls. Fernandez-de-las-Penas et al (2006) found that a FHP CV angle in standing of $47.9^\circ \pm 7.9^\circ$; $P < 0.001$ was associated with tension-type headache, and in an earlier study Fernandez-de-las-Penas et al (2005) found that patients with chronic tension-type headaches again had a smaller CV angle ($45.3^\circ \pm 7.6^\circ$ versus $54.1^\circ \pm 6.3^\circ$ for controls; $P < 0.001$). Tension-type headache is the most frequent form of headache in adults with a one-year prevalence of 38.3% for episodic tension-type headaches and 2.2% for chronic tension-type headaches. Watson and Trott (1993) found similar results (in relation to CV angle) for patients suffering from cervicogenic headaches; ($44.5^\circ \pm 5.5^\circ$ versus $49.1^\circ \pm 2.9^\circ$ for controls; $P < 0.001$).

Rating of Forward Head Posture
Therapists rate the severity of the anterior positioning of the head as minimal, moderate or maximal without any objective or numeric values (Yip et al, 2007). A decision regarding normality or otherwise is then based on a clinicians’ experience and perception of what constitutes a normal or “ideal” posture, and is therefore considered to be a potential source of error
as the decision is a subjective as opposed to objectively measured finding (Griegel-Morris et al, 1992).

One objective, and widely used method of assessing head posture is through measuring the craniovertebral (CV) angle (Watson and MacDonncha, 2000; Yip et al, 2007). The CV angle is the angle between a horizontal line through the spinous process of C7 and a line from the spinous process of C7 through the tragus of the ear (Figure 1). Other forms of objective measurement of FHP include various “head tilt angles” where the true horizontal is measured against angles used in conjunction with the tragus to the canthus (outer slit in the eyelid) (known as the Ear-Eye Line); the tragus to the bottom of the eye socket (Frankfurt Line); the tragus to the nasion (the middle of the naso-frontal suture); and the tragus to the infra-orbital notch (junction of the lateral third and medial two-thirds of the inferior orbital rim) (Ankrum and Nemeth, 2000; Moore and Dalley, 1999). Plain film radiographs of the head and neck to measure the relationships between bony structures, without reference to any external landmark have also been described (McAvinney et al, 2005; Peterson et al, 1997).

Figure 1. Measurement of cranio-vertebral angle.

Part Five: Reliability of Visual Postural Assessment

Having provided an overview of reliability, observation and visual assessment, skill acquisition, and posture, there will now be a summary of the research
literature related to the above topics. In total, eighteen studies shall be reviewed in order to introduce the topic and provide a contextual basis for the development of the research question of the present study.

Burt and Punnett (1999) evaluated the inter-rater reliability of postural assessment in a field study. Two observers independently evaluated 70 different job procedures in an automotive manufacturing facility. 18 different postures of the upper extremities and back were observed. Inter-rater agreement using kappa ($k$) ranged from $k = -0.08$ (95% confidence interval -0.14 – 0.03) for ulnar deviation – left, to $k = 0.47$ (95% confidence interval 0.28 – 0.67) for pinch grip – left. The authors concluded that inter-rater reliability was acceptable for some of the postural observations in their study. They suggest that inter-rater reliability of postural observations can be optimised when operational definitions are simple and unambiguous (ideally pre-tested); longer and multiple training sessions precede data collection; the number of postures observed is limited (and observations are prioritised); the level of detail is limited; and real-time observations are limited to jobs that do not involve rapid, dynamic movements. Longer observation periods and repeated observations may also improve the accuracy and precision of observational assessments. The use of only two observers is a weakness in this study as this limits the application of results to a wider population.

Watson and McDonncha (2000) assessed spinal views in order to grade the posture of subjects according to three categories; good posture, moderate defect, and severe defect. The sample size used was 114 males aged 15-17, randomly selected from two schools. Four photographic views were taken; anterior, posterior, lateral, and oblique. The subjects were shown a diagram or photograph of the required comfortable erect posture. The observers (number and training unknown) assigned a score of 5 (good posture), 3 (moderate deviation), or 1 (marked deviation) for each subject for 10 different parts of their body (e.g. ankle, knee, spinal lordosis and kyphosis, etc.). According to the authors, the results showed that the qualitative and quantitative approaches used in the study were “extremely reliable”, with the quantitative measures being slightly more reliable (90-100 quantitative, compared to 66-100 percentage agreement for the qualitative
method). However, these results are not a true indicator of reliability as the use of proportion of agreement does not account for chance agreement. Kappa is a measure of agreement that does account for chance. The formula to calculate kappa subtracts the proportion of agreement that could be expected by chance alone from the observed proportion of agreement, and can therefore avoid erroneous conclusions that agreement is good, when in fact it may simply be due to chance (Meeker and Escobar, 1998). This strongly appears to be the case in this study.

Fedorak et al (2003) investigated the reliability of visual assessment of cervical and lumbar lordosis. Six chiropractors, seven physical therapists, six physiatrists, four rheumatologists, and five orthopedic surgeons were recruited to evaluate the posture of photographed subjects. The subjects were a convenience sample of 36, 17 with back pain, and 19 without back pain. Kappa coefficients ($k$) were calculated to determine intra-rater and inter-rater reliability. Mean intra-rater reliability was $k = 0.50$ (95% confidence interval 0.20 – 0.98) and mean inter-rater reliability was $k = 0.16$ (95% confidence interval 0.00 – 0.48). No statistically significant difference existed among the five groups of clinicians or between the evaluation of the subjects with and without back pain. Fedorak et al concluded that intra-rater reliability of the visual assessment of cervical and lumbar lordosis was statistically ‘fair’, whereas inter-rater reliability was ‘poor’. These results were contrary to the hypothesis that visual assessment would have a high intra-rater and inter-rater reliability, and it was therefore stated that clinicians need to be made aware of the limitations of visual assessment, and that other tools, that were unstated, are more accurate and/or reliable, and should be used in combination with visual assessment to improve the quality of the spinal posture examination. Limitations in this study include the non-random selection of subjects and examiners, plus the small number of examiners. This study measured the consistent response by the examiner, and no analysis was done to evaluate how far the reliable measurements were from the true measurements (i.e. validity).

The study with both the subject and methods most closely related to this thesis was that of Ljubotenski (unpublished, 2006) who investigated the intra and inter-
rater reliability for visual assessment of the lumbar spinal curve and the influence that osteopathic clinical experience had on the reliability of the visual assessment. The study also investigated the effect of subject Body Mass Index (BMI) on the reliability of visual assessment. The study used a blinded test-retest design.

Twelve observers assessed 130 videos of subjects. The videos consisted of 60 recordings used twice and ten used three times to determine intra-rater and inter-rater reliability. Observers were three first, and three fifth year osteopathy students, and three osteopathic practitioners with less than five years experience, and three with fifteen or more year’s experience. The observer’s impression of the depth of a subject’s lumbar curve was recorded on a 100mm long visual analogue scale (VAS) with end markings representing ‘maximum lordosis (deep)’ and ‘minimum lordosis (flat)’. Inter-rater reliability was investigated using twelve randomly selected observers evaluating the lumbar curves of sixty subjects on separate occasions. The influence of increasing clinical experience was investigated by grouping the practitioners’ results in four groups according to clinical experience and comparing them. Interclass correlation coefficients (ICCs) were used for data analysis. Results showed that the intra-rater reliability for the four groups of observers combined was found to be ‘very large’ (according to the guidelines for interpretation of ICCs suggested by Hopkins (2000)) (ICC = 0.71; 90% confidence interval 0.47 – 0.86). The inter-rater reliability for the four groups of observers combined was found to be ‘large’ (ICC = 0.53; 90% confidence interval 0.29 -0.70). Junior students appeared to achieve lower intra-rater reliability compared to their more senior peers, and both the junior students and senior osteopaths appeared to have lower inter-rater reliability compared to the senior students and newly graduated osteopaths. To conclude, no meaningful differences in reliability were observed between the observer groups of different clinical experience. The internal validity of this study is weakened by the fact that all groups involved were either in the same year of the same programme, or graduated from the same programme at the same time except for the group of three osteopaths with fifteen years or more of experience. This last group is not homogeneous with the other groups, so any conclusions drawn when making comparisons between this group and others may be regarded as less valid.

Overall subject numbers in the study are low, plus all subjects were drawn from the Auckland population making generalisation to a greater population less valid.
Reliability of Visual Postural Assessment after Stroke

Carr et al (1999) investigated inter-rater reliability of postural observation after stroke. Six observers with varying degrees of experience in observing posture made observations of 19 aspects of the posture of a convenience sample of 57 stroke patients. The extent of inter-observer agreement was calculated using percentage agreement as well as kappa values. Results showed that acceptable percentage agreement, defined by Carr et al (1999) as being greater than 70%, for observations for the upper limbs was 67% (n=78) and 73% (n=55) for the lower limbs. In contrast, acceptable percentage agreement for observations relating to the head, neck and trunk was obtained for only 34% (n=50) of the results collected. Inter-observer agreement was not noticeably higher for the more experienced observers. Uneven distributions, that were not explained, reportedly made kappa values difficult to interpret and were therefore not reported. As with Watson and MacDonncha (2000), the authors relied on reporting their data through percentage agreement values. As previously mentioned these values are not a true indicator of reliability as the use of proportion of agreement does not account for chance agreement. Kappa is a measure of agreement that does account for chance. The poor statistical analysis used in this study means any conclusions drawn from the data are questionable.

Reliability of Visual Assessment of Gait

The intra- and inter-rater reliability of visual assessment procedures for gait analysis was investigated by Brunnekreef et al (2005). Observers included experienced, inexperienced and expert observers whose task was to individually evaluate video recorded gait patterns of subjects. The inexperienced observers were four students with no clinical experience of gait analysis. The group of experienced observers consisted of four senior physical therapists that had successfully completed a gait analysis course. The two expert observers were senior physical therapists with a vast knowledge and exceptional skills in gait analysis. The inter-rater reliability among experienced observers (ICC = 0.42; 95% confidence interval: 0.38 - 0.46) was comparable to that of the inexperienced observers (ICC = 0.40; 95% confidence interval: 0.36 – 0.44). The expert observers reached a higher inter-rater reliability level (ICC = 0.54; 95% confidence interval: 0.48 – 0.60). The average intra-rater reliability of the
experienced observers was 0.63 (ICCs ranging from 0.57 to 0.70). The inexperienced observers reached an average intra-rater reliability of 50.7 (ICCs ranging from 0.52 to 0.62). The two expert observers attained ICC values of 0.70 and 0.74 respectively. The authors concluded that clinical experience appeared to increase the reliability of visual gait analysis. However, the two expert observers in this study cooperatively developed the gait analysis form used in the actual study, which compromises the studies internal validity.

Another group of researchers, Wren et al (2005) investigated the reliability and validity of visual assessment of gait using a modified physician rating scale for crouch and foot contact. Thirty subjects with pathologic gait were evaluated “live” and using full- and slow-speed video. Observers (number and demographics unknown) assessed the subjects on various aspects of their gait. The fact that no information at all was provided regarding the observers basically renders the findings of this study useless. The authors concluded that visual assessment was valid and reliable for assessing only certain phases of gait, but such assessments are not an adequate substitute for computerised gait analysis testing due to their degree of inaccuracy.

The Influence of Experience on Reliability of Visual Assessment
Mann et al (1984) investigated intra- and inter-rater agreement among physical therapists of varying levels and types of experience for an assessment procedure involving palpation and observation that is advocated as part of the orthopaedic structural examination for determining the presence of an apparent leg-length difference. The subjects were three women and seven men who volunteered for the study. The eleven observers were three third year physical therapy students, three physical therapists with mean clinical experience of 3.5y ± 2.17, and five physical therapists with mean clinical experience of 5.3y ± 2.48. Correlation analysis was performed on the non-continuous nominal data. The assumption that physical therapists experienced in the technique of palpation and observation of iliac crest heights would show better agreement with this test than therapists with rehabilitation experience was not shown to be true. This finding indicates that the specific type of physical therapy experience may have little bearing on the
agreement of palpation and observation of iliac crests, but that familiarity with patient assessment can enhance the reliability of this test.

Somers et al (1997) investigated the influence of experience on the reliability of goniometric and visual measurement of forefoot position. Two clinicians with a minimum of ten years experience that routinely evaluate and treat dysfunction of the feet were recruited, and two physical therapy students were also recruited. The students underwent a training program seven days prior to the start of the study taught by a physical therapist with five years of experience in evaluating and treating foot dysfunction. Ten subjects (20-31 years old), free from pathology, were measured. Each foot was evaluated twice with the goniometer and twice with visual estimation by each tester. Intra-class correlation coefficients (ICCs) and coefficients of variation method error were used as estimates of reliability. There was no major difference in the intra-rater or inter-rater reliability between experienced and inexperienced testers, regardless of the evaluation used. Estimates of intra-rater reliability when using the goniometer, ranged from 0.08 to 0.78 for the experienced raters and from 0.16 to 0.65 for the inexperienced raters. When using visual estimation, values ranged from 0.51 to 0.76 for the experienced raters and 0.53 to 0.57 for the inexperienced raters. The authors concluded, that experience does not appear to influence forefoot position measurements, of the two evaluation techniques. However, the authors state that they have not addressed the issue of validity. All observers were trained to some degree in analysing foot dysfunction, yet all subjects had no dysfunction. This makes the subject group non-representative of what the observers would normally view in a clinical setting, thus blurring the distinction between observer groups and therefore any conclusions drawn from the data.

Validity of Visual Postural Assessment

The study by Bryan et al (1990) assessed the validity of physical therapists visual postural evaluation skills in assessing lumbar lordosis using photographs of clothed subjects compared to actual radiographic measurements. Forty-eight physical therapists rated subjects from least to greatest amount of lordosis. In 96 trials, only nine raters correctly rated the amounts of lordosis. A chi-squared test
showed a high correlation between ratings for two picture sets (i.e. high reliability); however, the observed rankings were not valid when compared with radiographic measures of lordosis. The results indicate low validity in assessing relative amounts of lumbar lordosis using photographs of clothed subjects. The question has to be asked of this study as to why the visual evaluation of posture was conducted using clothed subjects? For the purposes of evaluating a patient’s lumbar spine it would not be standard clinical procedure to have clothing over the area to be examined as obviously clothing covers the area thus making any judgment purely guesswork.

Reliability of Postural Assessment Involving the Use of Digitising Systems

With computers now everyday items, there has been integration of digital technology into the clinical setting. The following section provides reviews of the studies involving both computer digitisation of postural views and use of actual physical devices to measure posture. The overall aim of these studies was to determine if such tools are indeed more reliable (and therefore more valid) compared to the traditional method of visual observation, and as such may offer an alternative.

Braun and Amundson (1989) investigated the reliability of the use of a computer-assisted slide digitising system called the Postural Analysis Digitising System (PADS). The purpose of this study was to assess the within-day and between-day reliability of the PADS system to measure three aspects of head and shoulder posture. The authors state that clinical assessment of posture tends to be subjective in nature and “while improvement over time may be detected in a specific patient, it is difficult to compare patients to each other and to quantify the improvements”. This suggests that the authors feel subjective measures of analysis of head and shoulder posture are unreliable and they state; “previously cited methods of (postural) assessment do not report reliability testing”. Twenty male subjects were photographed in a neutral position, the maximally protracted position, and the maximally retracted position of the head and scapula. Ten subjects were evaluated once, and ten were evaluated twice. The slide photographs were analysed using PADS. Anatomical landmarks were used to determine angular relationships in the head, neck, and shoulders. The reliability
of the system was tested by calculating an intra-class correlation coefficient (ICC), student t-test, and the percent error for each position. The ICC values demonstrated a significant correlation between the measurements from two sessions for all positions (0.71 to 0.87), and overall it was concluded that the statistics showed that the three head positions were both reproducible and reliable. The methods and data analysis of this study are strong. The actual application of the PADS system into a clinical setting requires further investigation. This includes; costs of software and required hardware; practicality, such as those of time constraints, as well as the amount of actual benefit gained from the analysis in regards to aiding diagnosis and subsequent treatments.

Dunk et al (2004) investigated the reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. The purpose of the study was to determine the reliability of posture within subjects, as well as examine any differences in posture reliability between males and females. The subjects were fourteen (seven male, seven female) students. All had been free of low back pain during the previous six months. The subjects stood in a relaxed upright posture next to a vertical reference line and were required to attend three individual sessions, each consisting of three trials of three views of upright standing posture. Reflective fin markers were taped to the skin of each subject at the level of C7, T12, and L5. Digitisation of the images as well as calculation of angles was performed using customised software. Three-way analysis of variance (ANOVA) was performed for each angle in each view to examine any differences in angle. To evaluate intra-subject reliability across sessions, intra-class coefficients (ICCs) were calculated. Results revealed large coefficients of variance reflecting the substantial intra-subject variability, as well as poor to moderate agreement indicated by intra-class correlation coefficients (ICCs). Individual subject posture was poorly to moderately repeatable on a day-to-day basis. The authors state that the poor repeatability of postures documented by their study brings into question the validity of postural analysis for diagnostic use or tracking changes in response to treatment. There were methodological limitations in this study, the first being the use of only one observer. The observer could have been unusually good or bad from one examination to the next. Statistically, therefore, multiple observers are needed to average any artificially low or high intra-examiner data so as to be
more generalisable to the “normal” observer population. Second, their use of only 14 subjects is a questionable size for a reliability study.

Normand et al (2007) used a photographic digitiser (PosturePrint® system) to conduct their study: “Three-dimensional evaluation of posture in standing with PosturePrint: an intra- and inter-examiner reliability study”. The authors state “in today’s evidence based health care arena, it is unacceptable to evaluate patients with non-objective measures”. The authors believe that subjective measures of postural evaluation are not a valid or reliable form of evaluation. This is further evidenced by the fact that there is no mention of visual observation in regards to their extensive list of clinical tools available to evaluate posture. In this study three examiners performed delayed repeated postural measurements on forty subjects over two days. Each examiner palpated anatomical locations and placed 13 markers on the subjects before photography. On the digital photographs, using the computer mouse, examiners identified an additional 16 points. Using the PosturePrint® Internet computer system postures were calculated as rotations in degrees and translations in millimeters. For reliability, two different types (liberal and conservative) of inter- and intra-examiner correlation coefficients (ICC) were calculated. All the “liberal” ICCs were in the excellent range (>0.84). For the more “conservative” type ICCs, four inter-examiner ICCs were in the interval (0.5-0.6), 10 ICCs were in the interval (0.61-0.74), and the remainder were greater than 0.75. The authors concluded that this method of evaluating posture is reliable for clinical use. Again, as with the studies by Braun and Amundson (1989) and Dunk et al (2004), which utilise digitalisation for postural analysis, the actual practicality of using these systems in a clinical setting has to be called into question.

**Reliability of Postural Assessment Involving the Use of Measurement Devices**

Hart and Rose (1986) investigated the reliability of a method for the assessment of lumbar spine curvature. The study assessed the intra-rater reliability of a practitioner assessing the lumbar curvature of a patient using a ‘flexicurve’ ruler on different occasions. The mean intra-rater reliability obtained was ICC = 0.87 which is considered to be ‘very large’ (Hopkins, 2000) and indicates that a
practitioner using the flexicurve ruler can be very reliable at assessing the lumbar curvature of the same patient on multiple occasions. However, the authors’ validation was based on only six subjects meaning less generalisation to the broader population.

Youndas et al (1991) investigated reliability of measurements of cervical spine range of motion. Three methods were examined: use of a cervical-range-of-motion (CROM) instrument, use of a universal goniometer (UG), and visual estimation (VE). Measurements were made on 60 patients (39 women, 21 men. Age range 21 to 84 years), with orthopaedic disorders of the cervical spine, who were divided into three groups of 20 subjects each. The testers were 11 volunteer physical therapists. Their clinical experience ranged from 2 to 27 years. All subjects were tested in a standardised seated position using operationally defined goniometric placements and non-goniometric estimation techniques. Cervical flexion and extension, lateral flexion, and rotation were measured. Intra-class correlation coefficients (ICCs) were used to quantify intra and inter-tester reliability. Youndas et al (1991) state that as no universally acceptable levels had been adopted for correlation coefficients for the purpose of describing the reliability of measurements, the authors defined ICC values of 0.90 to 0.99 as high reliability, 0.80 to 0.89 as good reliability, 0.70 to 0.79 fair reliability, and 0.69 and below as poor reliability. The authors found that goniometric measurements of active-range-of-motion (AROM) of the cervical spine made by the same physical therapist had ICCs greater than 0.80 when made with the CROM device or the UG. When different physical therapists measured the same patient’s cervical AROM, the CROM device had ICCs greater than 0.80, whereas the UG and VE generally had ICCs less than 0.80. The results show that compared with goniometric techniques, the between-tester reliability of VE is poor. The authors state that before the study began a 60-minute training session using a written protocol that described each method of measurement was provided. It is not stated how much of this time was allocated to each of the three measurement methods. However, it does seem highly likely that more time was spent learning how to use a goniometer rather than the process of visual estimation, as the paper states that the goniometer had predefined placements that were outlined in detail in the paper, whereas nothing at all is stated about how the
observers were to achieve their visual estimation measurements. Had the same amount and detail of training been provided for visual estimation the results may have differed.

The study by Peterson et al (1997) investigated the validity and reliability of four objective techniques for measuring forward shoulder posture. Subjects were 25 males and 24 females. Subjects had a lateral cervical spine radiograph taken, from which the horizontal distance from the C7 spinous process to the anterior tip of the left acromion process was measured. Subjects then proceeded twice through a random order of four measurements: the Baylor square, the double square, the Sahrmann technique, and scapular position. These results were then used to determine the intra-rater reliability of each technique. Multiple regression analyses were performed on each measure’s mean scores to determine both the correlation with the predictive value for the radiographic measurement. The ICCs ranged from -0.33 to 0.77, and the coefficients of determination ranged from 0.10 to 0.59 (n=49). The researchers demonstrated clinical reliability for each technique; however, validity compared with the radiographic measurement could not be established. These techniques may have clinical value in objectively measuring change in a patient’s shoulder posture as a result of a treatment programme. The researchers state that before any of these measures could be universally recommended in clinical practice, future research is necessary to establish inter-rater reliability and assess each technique’s ability to detect postural changes over time.

Lundon et al (1998) investigated the inter- and intra-rater reliability in the measurement of kyphosis in postmenopausal women with osteoporosis. The two non-invasive measurement devices used were the DeBrunner’s kyphometer and the flexicurve ruler. The DeBrunner’s kyphometer is a device that closely resembles a protractor. These tools were used to measure the thoracic spinal kyphosis. Both methods achieved very large intra and inter-rater reliability, with ICC’s ranging between 0.81 and 1.

Arnold et al (2000) evaluated the reliability of posture measures for postmenopausal osteoporotic women. Twenty women were tested by the same
physical therapist on three visits and by two therapists on the last visit using: 1) height and forward head position; 2) carpenter’s trisquare; 3) flexicurve ruler; 4) New York Posture Rating Scale (plumbline reference used); and 5) posture classification where the therapist classified the subject’s posture as sway-back, flat-back, military type, kyphosis-lordosis, ideal or unable to classify (plumbline reference used as well as grid board). Intra-rater and inter-rater reliability for height, forward head position, and the flexicurve ruler were very good (ICC = 0.86-0.99), and low to good for the carpenter’s trisquare (ICC = 0.59-0.88), posture rating scale (ICC = 0.57 – 0.73) and posture classification (kappa = 0.53).

It was concluded that in osteoporotic women, standard measurement tools such as the flexicurve ruler are reliable, practical measures of posture. Observational rating scales may be useful, however inter-rater reliability is questionable. This current study will provide further investigation into how useful and reliable observation of posture is. The following section shall provide the rationale for the current study.

**Part Six: Rationale for Current Study**

The current study shall investigate the reliability of visual assessment of forward head posture in standing. The outcome of this study may help to determine the degree of emphasis placed upon visual observation of forward head posture in the clinical setting. If results show that reliability of observation of forward head posture is statistically significant then the process of visual observation may be regarded a worthwhile procedure. However, if results show visual assessment to be unreliable the question must be asked whether or not this routine clinical procedure should be adhered to or even adjusted in a manner to increase reliability, and corresponding usefulness in the clinical setting.

The only study identified in the literature that included osteopaths as observers in an investigation of visual observation of posture was that of Ljubotenski (unpublished, 2006). However, the sample size was only 12 observers. Fedorak et al (2003) provided a larger sample size of 28 observers and 36 subjects in the area of spinal posture assessment to confirm their finding of poor to fair reliability of visual assessment. The results of Fedorak et al (2003) contradict those of Ljubotenski (unpublished, 2006) who reported both high intra-rater and high
inter-rater reliability. Clearly more study in this field is required, which is the purpose of this current study that shall investigate the reliability of visual assessment of forward head posture in standing.

Section two of this thesis, the manuscript, presents the current study where laypeople, osteopathy students, and osteopathic practitioners shall be investigated to assess the effect that experience has on reliability of visual assessment as there is an expectation that more experienced practitioners have better observational skills.
REFERENCES


SECTION TWO: MANUSCRIPT

Note:
This manuscript has been prepared in accordance with the Instructions for Authors for the journal: Manual Therapy [See Appendix 13].

References to appendices (as outlined in square brackets) are to the Appendix of the dissertation and do not constitute part of the manuscript itself.
Reliability of Visual Assessment of Forward Head Posture in Standing

Author
A W Aitken
School of Health Science
Unitec New Zealand
Email: awaitken@gmail.com
Tel: + 64 3 455 8372
**ABSTRACT**

**Background.** Visual assessment of a patient’s standing posture is performed by osteopaths and other manual medicine practitioners as part of the physical examination. Assessment of forward head posture is an important part of this postural assessment. As it appears that a small number of studies have investigated the reliability of visual assessment of posture, it is hoped that this study will contribute further data on the reliability and consequent utility of this assessment method.

**Objective.** To investigate the intra and inter-rater reliability of visual assessment of forward head posture in standing and to determine the influence of the osteopathic clinical experience on reliability of visual assessment of forward head posture.

**Study Design.** Blinded test-retest design examining intra-rater and inter-rater reliability of the visual assessment of forward head posture of a subject in the standing position.

**Method.** Intra-rater reliability was investigated by blinding observers who visually rated a video clip of the forward head posture of the same subject (n=60) twice. Inter-rater reliability was investigated by comparing results of 78 randomly selected observers comprised of 16 laypeople, 40 osteopathic students, and 22 osteopathic practitioners. The influence of clinical experience was investigated by comparing observers from these various groups.

**Results.** Intra-rater reliability across groups was only slightly better than would be expected by chance alone. Inter-rater reliability was universally statistically poor to fair, with only first and second year osteopathic students providing a moderate level of reliability.

**Summary.** The results demonstrate poor statistical reliability of the visual assessment of forward head posture. More research is required in order to determine the variables that may confound statistical reliability of this commonly used clinical assessment technique.
INTRODUCTION

Osteopaths and other manual medicine practitioners often rely on visual assessment as part of the clinical examination of the musculoskeletal system. Together with information from history and physical examination, visual assessment informs the clinical reasoning process in order to reach a diagnosis.

The purpose of visual assessment is to gain information on visible defects, functional deficits and abnormalities of postural alignment (Magee, 2002). The assessment of the head and neck, specifically forward head posture, is an important part of this process as an anterior position of the head has been recognised as a factor that may contribute to the onset and perpetuation of cervical pain syndromes (Braun and Amundson, 1989). Therefore, information gathered during visual assessment may contribute to subsequent examination, diagnostic reasoning and treatment planning.

In manual medicine, reliability of all visual assessment is regarded as important, in patient assessment (Enwemeka et al, 1986; Refshauge et al, 1994). Intra-rater reliability is the measurement consistency of one examiner taking repeated measurements of a constant phenomenon at separate times (Meeker and Escobar, 1998). Inter-rater reliability is the consistency of measurements of the same phenomena between different examiners (Meeker and Escobar, 1998).

Although observation of static posture is routinely used by a range of manual medicine practitioners, there are only a small number of studies investigating the reliability of visual assessment. One study by Ljubotenski (unpublished, 2006) included osteopaths in the sample of observers being investigated. Here, the intra- and inter-rater reliability of visual assessment of the lumbar spine, and the influence that clinical experience had on the reliability of visual assessment, was examined. Intraclass correlation coefficients (ICC) were calculated to determine intra-rater and inter-rater reliability and results showed that intra-rater reliability for the 12 osteopathic observers was high (ICC = 0.71, 90% CI 0.47 – 0.86) in contrast to the inter-rater reliability, which was found to be moderate (ICC = 0.53,
90% CI 0.29 -0.70). No meaningful differences in reliability were observed between the observer groups of different clinical experience.

Fedorak et al (2003) investigated the reliability of visual assessment of cervical and lumbar lordosis. Clinicians (n=28) from five disciplines (chiropractors, physical therapists, physiatrists, rheumatologists, and orthopaedic surgeons) were recruited to evaluate photographic images of posture. The subjects were a convenience sample, with and without back pain. Kappa coefficients (k) were calculated to determine intra-rater and inter-rater reliability. Mean intra-rater reliability was \( k = 0.50 \) (95% CI 0.20 – 0.98) and mean inter-rater reliability was \( k = 0.16 \) (95% CI 0.00 – 0.48). No statistically significant difference existed among the five groups of clinicians or between the evaluation of the subjects with and without back pain. Fedorak et al concluded that intra-rater reliability of the visual assessment of cervical and lumbar lordosis was statistically ‘fair’, whereas inter-rater reliability was ‘poor’. It was therefore concluded that clinicians need to be made aware of the limitations of visual assessment and that other methods, presumably a measurement technique of known reliability, should be used in combination with visual assessment to improve the quality of the spinal posture examination.

Observation and visual assessment are widely understood and utilised components of the physical examination process (Magee, 2002; Seidel et al, 2003; Ward and American Osteopathic Association, 2003). Further research into this field is required in order to correct for and educate against the limitations in reliability this form of assessment presents.

The aim of this study was to i) investigate the intra-rater and inter-rater reliability of the visual assessment of forward head posture in standing, and ii) determine the influence of clinical experience on reliability of visual assessment.
METHODS

Study Design
A blinded repeated measures design was used to examine the intra-rater and inter-rater reliability for visual assessment of forward head posture in standing.

Participant Selection – Observers
There were three subgroups of observers: laypeople, osteopathy students, and osteopathic practitioners. Laypeople were recruited via word of mouth and poster advertising. In order for a layperson to be accepted into the study, they were required to have had no previous training or experience in a healthcare discipline. Osteopathy students were recruited from a university population using posters and word of mouth. Osteopathic practitioners were recruited using random sampling based on the public register of the Osteopathic Council of New Zealand. Random numbers (http://www.random.org) were used to select the observers. Batches of 30 CD-ROMs with accompanying cover letters [appendix 1,2] were posted to the randomly selected practitioners every week for seven weeks until the desired numbers of respondents was acquired.

Participant Selection – Subjects
Convenience sampling was used to recruit consenting subjects from the Unitec New Zealand Osteopathic Clinic patient population. In an attempt to enhance the extent to which subjects would be representative of patients that would normally be assessed in a clinical setting, a range of morphologies, gender and ages, with or without musculoskeletal symptoms were recruited for the study (n=60). Participants were enrolled in the study after providing written informed consent [appendix 3,4,5]. Details of subject’s age, weight, height and history of musculoskeletal and medical complaints were recorded. The Unitec Research Ethics Committee approved the study.

Inclusion and Exclusion Criteria
All observers were required to have access to an Internet enabled computer with a CD-ROM drive. Osteopathic practitioners were required to be currently registered with the Osteopathic Council of New Zealand. Osteopathic students
were required to be enrolled in the osteopathy degree programmes at the School of Health Science, Unitec New Zealand. Observers were excluded if a visual impairment or disability prohibited satisfactory viewing of the video clips. For the purpose of recording video, subjects were required to be able to stand stationary for 1 minute whilst disrobed from the waist up; understand written and spoken English and be aged 18 or above. Subjects were excluded if they had any distinguishing body features such as tattoos, birthmarks, scars etc. that may cause bias by enabling observers to recall the grade previously allocated to the subject.

Venue for Collection of Video
A well-lit plain coloured room with as few distinguishing fixtures as possible was established for the purpose of video recording of subjects at the Unitec New Zealand Osteopathic Clinic.

Materials and Procedures
A digital video camera (Panasonic NV-DS15) on a dolly-mounted tripod (SLIK 504 QF) raised to a height of 1500mm above ground level, was used to record subject posture [appendix 6]. Subjects disrobed to their underwear and a plain white towel was wrapped around the subjects’ waists to cover their underwear. To standardise positioning during video recording, subjects were instructed to stand “as stationary as possible with relaxed posture” on a wooden platform which had two heel-stops. The dolly was attached to the base of the platform via a 1.34m length of cord. Markers were fixed to the ground at 0° behind the subject and at a 120° left oblique view to mark the panning arc for the video recording. The camera was manually moved in a 1.34m arc over a duration of approximately 30 seconds, which was the time it took to guide the camera through the panning arc.

Image processing and compilation
The video footage was edited using iMovie 08 v.7.1.1 (Apple Computers Corp., CA, USA) on an Apple Macbook Pro® laptop computer. A GeeThree ‘Slick Transitions’ plug-in pack was used in iMovie to enable pixilation of the subjects’ faces for purposes of anonymity. The edited video clips (n=60) were then
compiled in randomised order into an auto-loading Adobe Flash® based CD-ROM computer programme specifically developed for this research project. Each of the 60 clips was repeated twice in random order to give a total of 120 clips.

Upon loading the CD-ROM, observers were presented with instructions regarding the required task [appendix 7,7a, 7b]. Observers entered a unique access code that was provided with each CD-ROM and were required to respond to the following preliminary questions using fixed-choice menus: 1) country they practiced in, 2) highest educational qualification 3) how many years they had either been practicing (or studying if students) and 4) if they were a practitioner, an additional question regarding how often they used visual inspection of posture was included. After answering these preliminary questions, the 120 video clip sequence commenced [appendix 8,9]. The observers responded to the video clips by sliding an animated horizontal sliding scale (via the computer mouse) to the position they thought best represented the subjects’ head posture. The scale was labelled from left to right, ‘maximal’, ‘moderate’ and ‘minimal’ (Figure 1). The observers were able to pause the movie at any desired angle, and could watch each clip an unlimited number of times if desired. If observers required a break they could stop the viewing session and re-enter their access code to continue from the subsequent un-rated clip [appendix 10]. Observers were encouraged to finish the task in one sitting if possible. Responses from observers were recorded in a secure online MySQL database for later data extraction and analysis. Data was exported from the database into a Microsoft® Excel 2004 for Mac® spreadsheet for offline analysis.
DATA ANALYSIS

Data Extraction
Raw data consisted of the observers’ unique access code (that allowed identification of the group and subsequent subgroup to which the observer belonged) as well as data for each of the observer’s rating of each video clip (n=120). These ratings were recorded as numerical values ranging from 0 to 100, where 0 represents ‘minimal’ forward head posture, and the value 100 represents ‘maximal’ forward head posture. The raw data was imported to Microsoft® Excel 2004 for Mac® and sorted into spreadsheets for the various groups: laypeople; students; practitioners; as well as the subgroups: Osteopathy students; year one, two, three, four, and five. 2) Practitioner experience; 0-4 years, 5-9 years, 10+ years. 3) Practitioner utility of observation; Always, Frequently, Regularly, Occasionally, and Never.

Data Normalisation
The CD-ROM computer programme used to record the observers judgement of a subjects forward head posture utilised a sliding scale that when moved recorded in the database a value from 1 to 100. These raw scores were not visible to observers. This continuous numerical data (0 to 100) was converted to normalised categorical data. This was achieved by subtracting an observers’ minimum rating from their maximum rating, thus giving the range over which the observer made their judgements. This range was then divided into thirds, representative of ‘maximal’, ‘moderate’, or ‘minimal’ forward head posture, which is how the majority of clinicians categorise their judgements of head posture (Yip et al, 2007).

For statistical weighting purposes (kappa is influenced by trait prevalence (distribution)) 21 of the 120 viewed video clips were used for the final data analysis. A left lateral view of each subject was captured from the video footage and the inclination of the head representative of ‘maximum’, ‘moderate’, and ‘minimal’ FHP was measured from a horizontal line to a line drawn between the tragus and C7 (cranio-vertebral angle (Figure 2)) using ImageJ version 1.37
software (Figure 3). This measurement technique from the tragus to C7 to the horizontal is a widely accepted procedure for measurement of forward head posture (Braun and Amundson, 1989; Refshauge et al, 1994; Watson and MacDonncha, 2000). A mean value was derived from two separate measurements. After graphing the results, seven subjects with ‘maximum’, ‘moderate’, and ‘minimal’ FHP were selected as the subjects whose view ratings were analysed (Figure 4).

**Calculation of Fleiss’ Kappa Coefficients (k)**

For the purpose of calculating categorical agreement using kappa (k) coefficients, an observation that was scored within the ‘maximal’ range was assigned the value ‘3’; ‘moderate’ was assigned the value ‘2’; and ‘minimal’ the value ‘1’. Using this procedure allowed data to be ‘normalised’ relative to the individual observer. For example, if an observers’ ‘maximal’ rating was 84 and ‘minimal’ was 15 on the sliding scale, the range would be 69 (84 – 15 = 69). This range of 69 was then divided into thirds. Any score less than the first third of this range was assigned the value ‘1’ representing ‘minimal’ forward head posture, any score greater than the first third of the range but less than the last third was assigned the value ‘2’ representing ‘moderate’ forward head posture, and a score greater than the last third of the range was assigned ‘3’ representing ‘maximal’ forward head posture. Fleiss’ kappa coefficients (k) were then calculated for general reliability between groups of raters using a 3 x 3 contingency table (Figure 5).

‘Calculating a Generalized Kappa Statistic for Use With Multiple Raters’ available for download from http://www.ccitonline.org/jking/homepage/kappa.xls was used to calculate the Fleiss’kappa coefficients and corresponding 95% confidence intervals.

In order to investigate if results had a statistically significant change due to the randomly selected set of seven subjects with ‘moderate’ forward head posture, two other randomly selected sets (n=7) of ‘moderates’ were compared. These sets (labelled set 1, 2 and 3 in Table 3 and Figure 7) showed no significant statistical difference.
Table 1. Interpretation of kappa (Landis and Koch, 1977).

<table>
<thead>
<tr>
<th>Range</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0</td>
<td>No agreement</td>
</tr>
<tr>
<td>0 – 0.19</td>
<td>Poor agreement</td>
</tr>
<tr>
<td>0.20 – 0.39</td>
<td>Fair agreement</td>
</tr>
<tr>
<td>0.40 – 0.59</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>0.60 – 0.79</td>
<td>Good agreement</td>
</tr>
<tr>
<td>0.80 – 0.99</td>
<td>Very good agreement</td>
</tr>
<tr>
<td>1.0</td>
<td>Perfect agreement</td>
</tr>
</tbody>
</table>

Calculation of Cohen’s Kappa Coefficients

Cohen’s kappa coefficients for pair-wise comparisons were calculated for intra-rater reliability for all observers in all groups and sub-groups [appendix 11]. As with the Fleiss’ kappa coefficients, the continuous data was converted to a ‘normalised’ categorical form. The observer rating for the first observation of a subject was compared to a second view of the same subject. As previously described with Fleiss’ kappa coefficient calculations, seven subjects with ‘maximum’, ‘moderate’ and ‘minimum’ FHP were used as the subjects whose posture was analysed. A weighted count was obtained (Figure 6). These values were analysed by using MedCalc® for Windows® v.9 statistical software. The interpretation of both Cohen’s and Fleiss’ kappa coefficients followed the description by (Landis and Koch, 1977) (Table 1).
RESULTS

**Characteristics of observers**
A total of 78 observers were recruited, comprising laypeople (n=16), osteopathy students (n=40) and osteopathic practitioners (n=22). The osteopathy students recruited for this study were stratified into subgroups representative of the year of study they were currently enrolled in; year 1 (n=7), year 2 (n=4), year 3 (n=8), year 4 (n=12), year 5 (n=9).

A total of 157 practitioners were contacted with 22 consenting to participation in the study. The clinical experience of practitioners ranged from 1 to 31 years. Practitioners were stratified into subgroups representing years of clinical experience (0-4 years (n=4), 5-9 years (n=4), and 10 plus years (n=14), and also into groups representative of the practitioners use of visual observation in their clinical practice; Always (n=8), Frequently (n=9), Regularly (n=4), Occasionally (n=0), Never (n=1). Laypeople were not stratified and remained as one group (n=16).

**Characteristics of subjects**
In total, 62 subjects were recruited for this study. The video clips of two subjects were excluded from the study due to the presence of tattoos. The final compilation of video clips was based on 60 subjects described in Table 2.

**Table 2. Subject characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Total (n=60)</th>
<th>Symptomatic (n=36)</th>
<th>Asymptomatic (n=24)</th>
<th>Male (n=42)</th>
<th>Female (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>37.9 (13.4)</td>
<td>41.4 (14.9)</td>
<td>32.7 (8.7)</td>
<td>40.4 (13.5)</td>
<td>31.8 (8.6)</td>
</tr>
<tr>
<td>Height (metres)</td>
<td>1.8 (0.1)</td>
<td>1.8 (0.1)</td>
<td>1.8 (0.1)</td>
<td>1.8 (0.1)</td>
<td>1.7 (0.1)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>74.9 (13.7)</td>
<td>74.4 (13.7)</td>
<td>75.7 (14.0)</td>
<td>82.1 (13.5)</td>
<td>65.7 (12.0)</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>24.4 (3.9)</td>
<td>24.3 (3.7)</td>
<td>24.4 (4.3)</td>
<td>25.4 (4.2)</td>
<td>22.7 (3.6)</td>
</tr>
</tbody>
</table>

Note: All figures are mean (SD)
Table 3 shows the results for intra-rater and inter-rater reliability by observer group. Intra-rater reliability, as measured by Cohen’s kappa, is presented as the median score for each observer group with the minimum and maximum range shown in brackets. The median standard error along with its’ range are presented in the adjacent column. The results from intra-rater reliability showed either ‘poor’ or ‘no agreement’ across all observer groups. Inter-rater reliability, as measured by Fleiss’ kappa, is presented as the mean score with a 95% confidence interval. The results from inter-rater reliability ranged from ‘poor’ (year three and five osteopathy students) to ‘moderate’ (year one and two osteopathy students). All other observer groups recorded ‘fair’ agreement. Figure 7 displays a graph of inter-rater reliability for all observer groups.
Table 3. Inter- and intra-rater reliability kappa values of observer groups

<table>
<thead>
<tr>
<th>Observer</th>
<th>Intra-rater Reliability (Cohen’s) kappa (Median)</th>
<th>Median Standard Error (Min and Max Kappa)</th>
<th>Interpretation</th>
<th>Inter-rater Reliability (Fleiss’) kappa (95% confidence interval)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Laypeople (moderate set 1) (n=16)</td>
<td>-0.035 (-0.182 - 0.184)</td>
<td>0.172 (0.128 – 0.215)</td>
<td>No Agreement</td>
<td>0.324 (0.271-0.378)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Laypeople (moderate set 2) (n=16)</td>
<td>-0.038 (-0.179 - 0.188)</td>
<td>0.175 (0.122 – 0.218)</td>
<td>No Agreement</td>
<td>0.398 (0.355-0.440)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Laypeople (moderate set 3) (n=16)</td>
<td>-0.036 (-0.181 - 0.186)</td>
<td>0.173 (0.122 – 0.215)</td>
<td>No Agreement</td>
<td>0.377 (0.336-0.418)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Students (moderate set 1) (n=40)</td>
<td>0.0285 (-0.273 – 0.406)</td>
<td>0.164 (0.115 – 0.195)</td>
<td>Poor</td>
<td>0.316 (0.301-0.331)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Students (moderate set 2) (n=40)</td>
<td>0.0294 (-0.281 – 0.411)</td>
<td>0.169 (0.118 – 0.199)</td>
<td>Poor</td>
<td>0.396 (0.379-0.412)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Students (moderate set 3) (n=40)</td>
<td>0.0290 (-0.0277 – 0.408)</td>
<td>0.171 (0.113 – 0.189)</td>
<td>Poor</td>
<td>0.384 (0.368-0.401)</td>
<td>Fair</td>
</tr>
<tr>
<td>Year 1 (n=7)</td>
<td>0.091 (-0.25 – 0.235)</td>
<td>0.154 (0.137 – 0.174)</td>
<td>Poor</td>
<td>0.417 (0.343-0.492)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Year 2 (n=4)</td>
<td>-0.086 (-0.206 – 0.031)</td>
<td>0.170 (0.154 – 0.179)</td>
<td>No Agreement</td>
<td>0.451 (0.306-0.595)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Year 3 (n=8)</td>
<td>0.034 (-0.161 – 0.406)</td>
<td>0.159 (0.125 – 0.176)</td>
<td>Poor</td>
<td>0.152 (0.079-0.224)</td>
<td>Poor</td>
</tr>
<tr>
<td>Year 4 (n=12)</td>
<td>0.052 (-0.273 – 0.381)</td>
<td>0.168 (0.147 – 0.195)</td>
<td>Poor</td>
<td>0.353 (0.310-0.396)</td>
<td>Fair</td>
</tr>
<tr>
<td>Year 5 (n=9)</td>
<td>0.021 (-0.243 – 0.237)</td>
<td>0.167 (0.115 – 0.183)</td>
<td>Poor</td>
<td>0.181 (0.125-0.237)</td>
<td>Poor</td>
</tr>
<tr>
<td>All Practitioners (moderate set 1) (n=22)</td>
<td>-0.0085 (-0.323 – 0.25)</td>
<td>0.164 (0.118 – 0.207)</td>
<td>No Agreement</td>
<td>0.266 (0.241-0.291)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Practitioners (moderate set 2) (n=22)</td>
<td>-0.0101 (-0.320 – 0.254)</td>
<td>0.165 (0.120 – 0.210)</td>
<td>No Agreement</td>
<td>0.332 (0.306-0.357)</td>
<td>Fair</td>
</tr>
<tr>
<td>All Practitioners (moderate set 3) (n=22)</td>
<td>0 to 4 yrs (n=4)</td>
<td>0.164 (0.117 – 0.209)</td>
<td>No Agreement</td>
<td>0.320 (0.293-0.346)</td>
<td>Fair</td>
</tr>
<tr>
<td>5 to 9 yrs (n=2)</td>
<td>-0.0555 (-0.323 – 0.23)</td>
<td>0.146 (0.118 – 0.170)</td>
<td>No Agreement</td>
<td>0.200 (0.047-0.352)</td>
<td>Fair</td>
</tr>
<tr>
<td>10 plus yrs (n=14)</td>
<td>0.106 (-0.092 – 0.236)</td>
<td>0.166 (0.157 – 0.172)</td>
<td>No Agreement</td>
<td>0.361 (0.233-0.490)</td>
<td>Fair</td>
</tr>
<tr>
<td>Always (n=8)</td>
<td>-0.031 (-0.256 – 0.25)</td>
<td>0.164 (0.131 – 0.207)</td>
<td>No Agreement</td>
<td>0.215 (0.179-0.251)</td>
<td>Fair</td>
</tr>
<tr>
<td>Frequently (n=9)</td>
<td>-0.029 (-0.184 – 0.165)</td>
<td>0.162 (0.141 – 0.177)</td>
<td>No Agreement</td>
<td>0.220 (0.157-0.283)</td>
<td>Fair</td>
</tr>
<tr>
<td>Regularly (n=4)</td>
<td>0.000 (-0.323 – 0.25)</td>
<td>0.165 (0.118 – 0.207)</td>
<td>No Agreement</td>
<td>0.314 (0.257-0.370)</td>
<td>Fair</td>
</tr>
<tr>
<td>Occasionally (n=0)</td>
<td>-0.016 (-0.142 – 0.103)</td>
<td>0.154 (0.131 – 0.171)</td>
<td>No Agreement</td>
<td>0.244 (0.102-0.385)</td>
<td>Fair</td>
</tr>
<tr>
<td>Never (n=1)</td>
<td>-0.073 (-0.073 - -0.073)</td>
<td>0.159 (0.159 – 0.159)</td>
<td>No Agreement</td>
<td>(n=1, so no result)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 7. Inter-rater reliability by observer group
DISCUSSION

Overview
This study was conducted to examine the reliability of a clinical assessment technique commonly used by osteopaths and other manual medicine practitioners to assess forward head posture. To our knowledge the reliability of assessing a subjects’ forward head posture has not been previously reported within the scientific literature, therefore, direct comparison of results to previous studies is not possible. Intra-rater reliability across groups was found to be only slightly better than expected by chance. Inter-rater reliability was, on average, statistically ‘poor’ to ‘fair’, with only first and second year osteopathy students demonstrating a ‘moderate’ level of reliability. Visual observation in the examination procedure was found to be statistically poor, independent of practitioner experience or procedure use. The reliability of visual observation does not increase with practitioner utilisation of the visual examination procedure. Both inter and intra-rater reliability for the visual assessment of forward head posture was less than that which is accepted as being clinically useful ($k \geq 0.4$) (Fjellner et al, 1999).

It is unclear why first and second year osteopathy students out-performed practitioners, however, one reason may be that these students have more recently learnt a particular method and consciously practice the skill of visual assessment of posture whereas other groups of observers, such as practitioners and more senior students, may have developed their own unique routine that may differ from others and also may not be as regularly practiced. It was also hypothesised that the novice first and second year students may have been more careful in their judgements, and therefore may take more time to make their decision. A sub-analysis was conducted that indicates that the duration of response times were very consistent from the first to last view across all observer groups; first and second year students did not spend more time making their judgements [appendix 12].

The more senior students may place less emphasis upon posture in their approach to patient examination and therefore don’t rehearse the visual assessment of
posture as much. The practitioners are from a varied number of different osteopathic training institutes, each with varied training, and each with a different emphasis on visual assessment of posture. The practitioners also differ in the length of time since their training, so even if they were educated in the same osteopathic educational institution their way of doing things will evolve to their own needs over time.

Experience does not appear to be a determinant of reliability. This is now something that has been observed in other studies; across practitioner groups with similar clinical experience (Fedorak et al, 2003) and also across different years of experience with osteopaths (Ljubotenski, unpublished, 2006), as well as different years of experience with physical therapists, (Brunnekreef et al, 2005; Somers et al, 1997).

Palpation, which is another form of clinical assessment, has also shown poor correlation between experience and reliability. Two studies involving both observation as well as palpation of anatomical landmarks reported negligible differences in reliability between experienced practitioners and students. Kmita and Lucas (2008) investigated the intra- and inter-examiner reliability of physical examination involving palpation and observation of the pelvis. Results showed that two experienced osteopaths and two final year osteopathic students had low inter-examiner reliability. The median observed percentage agreement between examiners for each anatomical landmark ranged from 33 to 50%. The intra-examiner reliability ranged from ‘less than chance’ to ‘perfect’ agreement. These broad findings led the authors to state that intra-examiner reliability has limited value as it only assesses how well an examiner agrees with their prior findings, and does not provide information about test reproducibility. In a similar earlier study, Mann et al (1984) tested agreement for palpation and observation of iliac crest heights using eleven physical therapists of various experience levels. Results showed that experienced physical therapists had only slightly higher intrarater and interrater agreement than student physical therapists.

Although there are no studies that are directly comparable to the current study, four authors have investigated the reliability of posture. In a recent similar study,
Ljubotenski (unpublished, 2006), who investigated visual assessment of the lordotic lumbar spinal curve, found that intra-rater reliability was ‘very large’ (ICC = 0.71) and inter-rater reliability was ‘large’ (ICC = 0.53). No meaningful differences in reliability were observed between the observer groups of different osteopathic clinical experience. The modest number of observers (n=12) limits the extent to which the findings can be generalised to a broader population of practitioners. The current study aimed to address the issue of small sample size and resultant lack of generalisability.

The study most comparable to the present study is that of Fedorak et al (2003) who investigated the reliability of visual assessment of cervical and lumbar lordosis between various practitioner groups. As with the current study, a blinded test-retest design was employed and kappa coefficients used to analyse observer reliability. Fedorak et al concluded that intra-rater reliability was ‘fair’, whereas inter-rater reliability was ‘poor”. The overall conclusion was that visual assessment of cervical and lumbar lordosis was not reliable which echoes the results from the current study. One criticism of the study is that the researchers did not employ random sampling of the patients or practitioners meaning that results may not be truly representative of the population they are supposed to represent. The present study employed random sampling.

In another study, Watson and MacDonncha (2000) assessed overall spinal views in order to grade posture of 114 subjects according to categories of good, normal, or bad posture, and reported more than 60% agreement on any given comparison between multiple assessments of the same subject. However, the quality of reporting in this study was poor. It was not stated how many observers did the rating; therefore, the extent to which we can generalise these results is confounded because the training of the observers is not apparent. Also, the use of proportion of agreement in this study does not account for chance agreement and is not considered the preferential mode of analysis. The appropriate statistic would be kappa coefficients for categorical agreement, or Intra-class Correlation Coefficients for continuous data (Meeker and Escobar, 1998). Kappa is a measure of agreement that does account for chance. The formula to calculate kappa subtracts the proportion of agreement that could be expected by chance alone.
from the observed proportion of agreement, and can therefore avoid erroneous conclusions that agreement is good, when in fact it may simply be due to chance (Meeker and Escobar, 1998).

Carr et al (1999) investigated inter-rater reliability of postural observation after stroke, the results for the six observers showed acceptable percentage agreement for observations of posture of the upper limbs in 67% of the observations, but only 34% agreement for observations relating to the head, neck and trunk. Similar to Watson (2000), the use of percentage agreement provides a less than ideal measure of reliability. This result from Carr et al (1999) along with that of Ljubotenski (unpublished, 2006) raises the question as to whether or not observation of the head and neck is more difficult to evaluate compared with that of the lower back and trunk. It is proposed that this is perhaps due to the physically smaller size of the cervical spine when compared to the lumbar spine, making any difference harder to judge.

The current study and previous literature in this field indicate that reliability of visual observation of posture is modest. It appears that the lumbar spine may be easier to judge from an observational perspective.

In this study the inter-rater reliability was calculated using the average of view one and view two (of the same subject) which removes the influence of high and low values. This approach was used because an average rating is arguably a more representative rating of what the true rating for that subject might be. To illustrate this point, for a task requiring judgement of multiple views of the same item, an average of the ratings will clearly provide a better representation of the views rather than just one rating. This average rating also lessens the effect of any covert cues that may trigger the observer’s memory when viewing a subject for the second time.

Strengths of the current study include; it evaluated a spectrum of symptomatic subjects who were recruited from a clinical setting and the observations were performed by observers representative of those who would normally perform the test in practice. Therefore results from this study are applicable to the wider
population of practitioners of similar training that use observational procedures in their clinical assessment.

The reference standard for postural assessment of forward head posture is a measured cranio-vertebral (CV) angle (Figure 2). In our study observers were blinded to this reference standard data. Observers were blinded to the findings of other observers in the study, as well as their own prior ratings, and clinical information about the subjects was not available and therefore could not influence the study outcome. In studies of this nature, additional cues such as tattoos, scars, birthmarks etc. may influence observers’ judgements particularly for the second view as they may enable a rater to recall this distinctive feature. For this reason subjects with unique identifying features were excluded from the study. Computer pixilation of subjects’ faces was used to prevent identification of subjects that may have provided a possible memory cue for subsequent ratings, and to provide subject anonymity. Finally, the order in which observers examined the subjects was randomised thus providing an impartial sample and consequently a valid basis for statistical inference.

The process of watching the 120 video clips took approximately 45 minutes, so it is possible that viewer fatigue may have influenced the results. A sub-analysis was therefore conducted that indicated that duration of response times were very consistent from the first to last view thus indicating that observers did not rush through the video clips leading to a potential cursory examination of the video clips [appendix 12].

In order to ensure that the random selection of the seven subjects used to represent the ‘moderate’ forward head posture group did not affect the results, the analysis was repeated with three different randomly selected groups of ‘moderates’. These three groups were consistent with overlapping confidence intervals; therefore we are confident that the random selection had no influence on the outcome of the results.

The sample of practitioners (n=22) in the current study represented 6.3% of osteopathic practitioners currently holding annual practicing certificates in New
Zealand. However, the extent to which the results are applicable outside of New Zealand is unclear. Applicability to the NZ student population may be more generalisable when compared to the practitioner group as the student sample (n=40) represented 27% of the osteopathic student population in New Zealand. Results may be of interest to those outside of New Zealand, as anecdotally students in Australia and the United Kingdom appear to have broadly similar pre-professional training to students in New Zealand.

The main limitation of this study was the use of a two-dimensional medium (video footage) to make judgements upon a three-dimensional object (the subject’s posture). The use of video in this study ensured that the trait under review, forward head posture, was stable between views. However, the extent to which videos may fully represent real life is less clear. There appear to be no studies in any field that have compared the use of real subjects with those of videoed subjects to determine the representativeness of video. Future research could attempt to address this issue by comparing the results of a real-time subject versus the very same subjects video-footage (filmed at the same time as the real-time evaluation).

There isn’t a universally accepted grading scheme for the assessment of head and neck posture. Therefore in this study we generated the subjective grading scheme of ‘maximum’, ‘moderate’ and ‘minimal’ to represent a subject’s degree of FHP. In so doing we may have introduced a grading scheme that the observers were unfamiliar with, so a future study could address this by investigating the effect of training in visual assessment to determine if it improves the reliability of visual assessment of FHP. Future research could also look at the validity of visual assessment of forward head posture in order to see how far the measurements were from the true measurements, as well as investigate to what extent practitioners use observation of posture in their clinical reasoning process.
CONCLUSIONS

Visual assessment of head and neck posture is a frequently utilised clinical procedure used by osteopaths and other manual medicine practitioners within the broader context of patient examination and assessment. The results of this study show that visual assessment demonstrates poor statistical reliability and brings into question the clinical utility of the procedure.

As visual assessment of patient head-neck posture is one of several clinical examination procedures undertaken, it is unknown how such individual examination components contribute *en masse* to a wider picture of a patient’s clinical assessment. Whilst individual clinical procedures may lack statistical reliability, it may be that when considered together they possess better overall statistical reliability. This is a matter for further investigation. It is therefore the recommendation not to use visual assessment alone to examine a patient’s head posture, especially with regard to the formulation of a diagnosis or treatment plan.
REFERENCES


Figure 1. Screenshot showing the ‘sliding scale’ computer interface observers used in order to assess a subjects’ degree of forward head posture, as well as the play/pause interface and close and save button.
Figure 2. Measurement of cranio-vertebral angle.
12a. Subject used as one of a group of seven to represent maximal FHP (angle of 28.23°)

12b. Subject used as one of a group of seven to represent moderate FHP (angle of 50.98°)

12c. Subject used as one of a group of seven to represent minimal FHP (angle of 69.78°)

Figure 3. Example measurements of FHP using ImageJ v.1.37.
Figure 4. Graph showing selection of subjects for kappa statistics.
### Calculating a Generalized Kappa Statistic for Use With Multiple Raters

Calculations based on equations presented in Fleiss (Statistical Methods for Rates and Proportions, 1981, pp. 229-232)

Enter # of raters (m): 16
Enter # of subjects (n): 21
# of categories (k): 3

For each item below, enter the number of raters who placed the item into each respective category (delete/add rows as necessary):

<table>
<thead>
<tr>
<th>n=21</th>
<th>Subject ID</th>
<th>Max</th>
<th>Moderate</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max (n=7)</td>
<td>14</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>2</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>2</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>0</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Moderate (normal) (n=7) (set 1)</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>2</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>2</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>0</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Min (n=7)</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>4</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>2</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>1</td>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

***************

**OVERALL**

\[
\begin{align*}
\text{num} &= n_{\text{set1}} \times n_{\text{set2}} \times \text{m} = 1352 \\
p(e) &= 0.651 \\
SE_{\text{overall}} &= 0.087 \\
z &= 2.657 \\
p\text{-value} &= 0.007875 \\
CI_{\text{lower}} &= 0.061 \\
CI_{\text{upper}} &= 0.404
\end{align*}
\]

**Figure 5.** Example calculation of Fleiss’ kappa 3 x 3 contingency table for laypeople.
**Layperson 007**

<table>
<thead>
<tr>
<th>View 1</th>
<th>View 2</th>
<th>Count (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 6.** Example weighted count for Cohen’s kappa calculation.
APPENDICES

Appendix 1. Example cover letter for post-out to osteopathic practitioners.
Appendix 2. CD-ROM cover.
Appendix 3. Participant consent form (subjects).
Appendix 4. Participant information sheet (subjects).
Appendix 5. Participant information sheet (observers).
Appendix 6. Photographs of platform, video camera and dolly system.
Appendix 7. Screenshots showing sequence of CD-ROM information sheet and drop-box questionnaire.
   7a. Example of entering information as a student observer.
   7b. Example of entering information as a practitioner.
Appendix 8. Example of video sequence of subject from CD-ROM.
Appendix 9. Freeze frames of video footage of subjects with varying degrees of forward head posture.
Appendix 10. Example of save and close option.
Appendix 11. Cohen’s kappa (linear weighted) inter-rater statistics for each individual observer.
Appendix 12. Response time versus subject number graph.
Appendix 13. The instructions for authors for the journal: Manual Therapy.
Appendix 1. Example cover letter for post-out to osteopathic practitioners.

Date

Practitioners Name
Practitioners Address

Dear Practitioner

My name is Andrew Aitken. I am a final year osteopathy student from Unitec in Auckland. I would like to invite you to participate in a research project involving assessment of a patient's head and neck posture. We are seeking to recruit osteopaths and osteopathy students to participate in this research. As a registered osteopath, your name was randomly selected from the public register of osteopaths.

Your participation will provide us with valuable information about how clinical observers make judgements about head and neck posture. You'll be asked to watch a series of short video clips that show a range of different head and neck postures. You'll need to use a computer to run the application that comes on the CD-ROM enclosed with this letter. The software will play automatically when placed in the CD drive of any PC computer (sorry, it won't play on a Macintosh). You'll need to enter the code printed on the front of the CD (also located on the case) for the software to run. There are further instructions available as you start the software. Please note that you'll need to be connected to the Internet while you undertake the experiment, as this is how we collect the data.

The experiment should take approximately forty-five minutes to one hour to complete, and can be paused and resumed at any time by simply re-entering your access code. To hasten analysis of the subjects' head posture you can simply press the pause button and then drag the video play-head to your desired view of the subject. A 'thank you' message will appear on screen to let you know when the assessments are complete.

I realise your time is valuable and thank you for considering participation in this project. If you have any questions please do not hesitate to contact me.

Yours sincerely

Andrew Aitken
E: awaitken@gmail.com    t: (09) 627 4191    m: 021 066 5577

P.S. If you do not wish to participate I do not need to be notified, nor do I need the disc to be returned.
Appendix 2. CD-ROM cover.

Information and Concerns
If at anytime you have further questions about the research project you may contact Andrew Atkin at: email awwakien@gmail.com, telephone (09) 274 1191, or mobile 021 088 5587.

Researchers
Andrew Atkin, BSc, BPhEd, PCDOp, SportsMed, BA(Human Biology)
This project is being supervised by:
Robert Morgan, BSc, BSc(Ch)Ed, MHSc(Osteopathy)
Chris McGrath DO (UK), PostGradBiol, PostGradOsteoHealth, MSc(Anatomy), PhD

Unitec
NEW ZEALAND
Te Whakatiro a Ora

Observation of Head Posture
Masters Research Project
CD-ROM

Instructions:
Connect computer to internet.
Place disc into CD drive.
The program will autorun after a few seconds.
Enter the Access Code: __________________________, at the prompt.
Follow instructions.

Note: Full instructions are supplied on the CD-ROM.
Appendix 3. Participant consent form (subjects).

Clinical Observation of Head Posture Assessment

Consent Form (Subjects)

This research project investigates how osteopathy students and practitioners make observations of head and neck posture. The research is being conducted by Andrew Aitken from Unitec New Zealand, and will be supervised by Robert Moran and Chris McGrath.

Name of Participant: ……………………………………………………………………

I have seen the Information Sheet dated ………………………………… for people taking part in the Clinical Observation of Head Posture Assessment Project. I have had the opportunity to read the contents of the information sheet and to discuss the project with Andrew Aitken and I am satisfied with the explanations I have been given. I understand that taking part in this project is voluntary (my choice) and that I may withdraw from the project at any time up until the stage of data analysis (approximately July 2007) and this will in no way affect my access to the services provided by the Unitec New Zealand Osteopathic Clinic service or any other support service.

I understand that I can withdraw from the video recording if, for any reason, I want this.

I understand that my participation in this project will be confidential and that no material that could identify me will be used in any reports on this project.

I understand that the video clips shall be used in a computer presentation which will be archived and might be used by other researchers with the approval of the Unitec Ethics Committee. I give consent for the video clip of my standing posture to be used for further research in the knowledge that the research purposes will be similar to this project, and that my identity will not be revealed at any stage.

I have had enough time to consider whether I want to take part.

I know whom to contact if I have any questions or concerns about the project.

The principal researcher for this project is Andrew Aitken email: awaitken@gmail.com phone: (09) 828 5750, mobile: 021 066 5577.

Signature………………………………………………   Participant ……… (date) DOB:

Project explained by………………………………………………………… Height:

Signature…………………………………………………. ………..(date) Weight:

The participant should retain a copy of this consent form.

UREC REGISTRATION NUMBER: (680.2007)
This study has been approved by the UNITEC Research Ethics Committee from 28 March 2007 to 31 December 2009. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Committee through the UREC Secretary (Tel: + 64 9 815-4321 ext 7248 or by email ethics@unitec.ac.nz). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.
Appendix 4. Participant information sheet (subjects).

PARTICIPANT INFORMATION SHEET (SUBJECTS)

Clinical Observation of Head Posture Assessment

About this Research
You are invited to take part in a research project being conducted by Andrew Aitken, a Master of Osteopathy student from Unitec New Zealand. This research project examines how consistently practitioners and osteopathy students are in making observations about posture.

Your participation will provide us with a video footage (a ‘virtual patient’) that we can use in a study that will involve osteopathy practitioners and students. A video sequence will be taken in one of the treatment rooms at the Unitec New Zealand Osteopathic Clinic. If you choose to participate you’ll be required remove your outer clothes so that we can observe from the top of your head to above your waist. You’ll be provided with a towel to wear around your waist in order to cover your underwear. Once you have assumed the desired position you’ll be asked to hold still for approximately 20 seconds while the researcher makes a recording using a video camera. The whole procedure will take approximately 5 minutes of your time. I can show you a short video clip as an example of the type of recording that I require for this research project.

Researchers
Andrew Aitken, BSc, BPhed, PGDpSportsMed, BAppSc(Human Biology). This project is being supervised by: Robert Moran, MHSc (Osteopathy) School of Health Science, Unitec. and Chris McGrath DO (UK), PostGradBiomech, PostGradOccHealthPrac, MSc (Anatomy) School of Health Science, Unitec.

Information and Concerns
If at anytime you have further questions about the research project you may contact Andrew Aitken at: email awaitken@gmail.com, telephone (09) 828 5750, or mobile 021 066 5577.

Privacy
Identity of participants will be protected in the following ways:

- The video sequence will include a view from the top of the head down, filmed from the side only – we don’t record any front on video
- Your identity will be disguised by either use of a face mask or video editing software to ‘blur’ out facial features.
- Participants with distinctive features such as tattoos, birth marks, scars, etc. aren’t eligible for inclusion in the research as these may be identifiable features.
- The participants underwear will be covered by a white towel wrapped around their waist.
- The data collected by the researcher will not include the participants’ name, address, phone numbers, etc. Only the participants’ gender, age, height and weight will be recorded.
- Any video images collected during the research process will be stored on a CD-ROM, DVD or video tape, and will be secured by a password or stored in a lockable cabinet in order to ensure that no unauthorised persons gains access to it.

A copy of the final draft of the final research project will be made available at the Unitec New Zealand library. All participants are welcome to view this. We would like to be able to use the video clips that we collect for other similar research projects in future (with the approval of an Ethics Committee). Any clips that are used in future will be processed to ensure your identity is protected.

Finally we would like to thank you for your valuable contribution to this research project.

UREC REGISTRATION NUMBER: (680.2007)
This study has been approved by the Unitec Research Ethics Committee from 28 March 2007 to 31 December 2007. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Committee through the UREC Secretary (Tel: + 64 9 815-4321 ext 7248 or by email ethics@unitec.ac.nz). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Thank you.
Appendix 5. Participant information sheet (observers).

INFORMATION SHEET

Observation of Head Posture

About this Research
You are invited to take part in a research project conducted by Andrew Aitken, a Master of Osteopathy student from Unitec New Zealand.

Your participation will provide us with valuable information about how observers make judgements about head and neck posture. You’ll be asked to watch a series of short video clips that show a range of different standing postures. You’ll need to use a computer to run the application that comes on the CD-ROM you have been provided with. The software should play automatically when placed in the CD drive of any PC computer (sorry, it won’t play on a Macintosh). If it doesn’t autostart after 30 seconds then double click on “VideoAnalysis” icon.

You’ll need to enter the code printed on the front of the CD (also located on the case) for the software to run. There are further instructions available as you start the software. IMPORTANT: YOU WILL NEED TO BE CONNECTED TO THE INTERNET WHILE YOU UNDERTAKE THE EXPERIMENT, as this is how the data is collected. You may experience difficulties running the CD using a networked computer that restricts administrative access to users.

What you need to do

Your task is to judge the ‘forward head posture’ of those people you see on the video clips using a scale of ‘maximal’, ‘moderate’, and ‘minimal’. The slider is to be moved to the position on the scale that you think best represents the position of head posture of the subject in the video clip. The video presentation will take around 45 minutes to an hour to complete. It is preferable that you complete the viewing of all clips in one session if you can but if necessary you can stop at any time and resume at a later time (the software automatically returns to the correct position). To hasten analysis of the subjects’ head posture you can simply press the pause button and then drag the video play-head to your desired view of the subject. A ‘thank you’ message will appear on screen to let you know when the assessments are complete.

Demonstration screen-shot outlining instructions
Consent
There is no consent form for this project. Your consent to participate in this project is implied by playing the software and undertaking the task. If you don’t wish to participate you need simply not run the software or complete the task. Taking a CD does not obligate you to participate.

Privacy
Identity of participants will be protected in the following ways:

- The results presented in any report will be pooled from all participants.
- The program asks you to enter some details about your osteopathy training and years of clinical experience but YOUR NAME IS NOT RECORDED.
- Any data collected during the research process will be stored on a password protected computer or in a lockable cabinet in order to ensure that no unauthorised persons gain access to it.

A copy of the research project report will be made available at the Unitec New Zealand library. All participants are welcome to view this or a copy can be sent to you directly if you email Andrew Aitken at the address below.

Finally we would like to thank you for your valuable contribution to this research project.

Information and Concerns
If at anytime you have further questions about the research project you may contact Andrew Aitken at: email awaitken@gmail.com, telephone (09) 627 4191, or mobile 021 066 5577.

Researchers
Andrew Aitken, BSc, BPhEd, PGDipSportsMed, BAAppSc(Human Biology)

This project is being supervised by:
Robert Moran, Bsc, BSc (ClinSci), MHSc (Osteopathy)
Chris McGrath, DO (UK), PostGradBiomech, PostGradOccHealthPrac, MSc (Anatomy), PhD

Thank You.

UREC REGISTRATION NUMBER: (680.2007)
This study has been approved by the UNITEC Research Ethics Committee from 28 March 2007 to 31 December 2009. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Committee through the UREC Secretary (Tel: + 64 9 815-4321 ext 7248 or by email ethics@unitec.ac.nz). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Note: A hardcopy of this form can be found in the CD cover.
Appendix 6. Photographs of platform, video camera and dolly system.
Appendix 7. Screenshots showing sequence of CD-ROM information sheet and drop-box questionnaire.
Consent
There is no consent form for this project. Your consent to participate in this project is implied by playing the software and undertaking the task. If you don’t wish to participate you need simply not run the software or complete the task. Taking a CD does not obligate

Privacy
Identity of participants will be protected in the following ways:

- The results presented in any report will be pooled from all participants.
- The program asks you to enter some details about your osteopathy training and years of clinical experience but YOUR NAME IS NOT RECORDED.
- Any data collected during the research process will be stored on a password protected computer or in a lockable cabinet in order to ensure that no unauthorised persons gain access to it.

A copy of the research project report will be made available at the Unitec New Zealand library. All participants are welcome to view this or a copy can be sent to you directly if you email Andrew Aitken at the address below.

Finally, we would like to thank you for your valuable contribution to this research project.

Information and Concerns
If at anytime you have further questions about the research project you may contact Andrew Aitken at: email awa1tken@gmail.com, telephone (09) 627 4191, or mobile 021 066 5577.

Researchers
Andrew Aitken, BSc, BPHEd, PGDipSportsMed, BAppSc(Human Biology).
A copy of the research project report will be made available at the Unitec New Zealand library. All participants are welcome to view this or a copy can be sent to you directly if you email Andrew Aitken at the address below.

Finally, we would like to thank you for your valuable contribution to this research project.

**Information and Concerns**

If at anytime you have further questions about the research project you may contact Andrew Aitken at: email aaitken@gmail.com, telephone (09) 627 4191, or mobile 021 666 5577.

**Researchers**

Andrew Aitken, BSc, BPhilEd, PGDipSportsMed, BA(Hons)(Human Biology).

This project is being supervised by:

Robert Moran, BSc, BSc(ClinSci), MMed(Osteopathy)

Chris McGrath DO (UK), PostGrad(Sketch), PostGradOccHealthPrc, MSc (Anatomy), PhD

**Thank You.**

**UREC REGISTRATION NUMBER:** (680.2007)
This study has been approved by the Unitec Research Ethics Committee from 28 March 2007 to 31 December 2009. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Committee through the UREC Secretary (Tel: +64 9 815-4321 ext 7248 or by email ethics@unitec.ac.nz). Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Note: A hardcopy of this form can be found in the CD cover.

---

Please enter your unique access code located on the case which this CD came in.

[DemoAccessCode]

Submit
Please fill out all fields below. If you wish to indicate an option not available from the drop-down list, select the last item on the list and enter a custom answer in the text box which appears.

Which country do you practice in?

Please Select...

Please indicate the highest educational level in ANY area of study.

Please Select...

Please select your status.

Please Select...
Please fill out all fields below. If you wish to indicate an option not available from the drop-down list, select the last item on the list and enter a custom answer in the text box which appears.

Which country do you practice in?

New Zealand

Please indicate the highest educational level in ANY area of study.

- Please Select...
- Diploma
- Bachelors Degree
- Honours Degree
- Masters Degree
- Doctoral Degree

Please select your status.

- Please Select...
- Please Select...
- Student (not yet registered to practice)
- Practitioner (Registered/Licensed if applicable)
Appendix 7a. Example of entering information as a student observer.

Appendix 7b. Example of entering information as a practitioner.
Please fill out all fields below. If you wish to indicate
an option not available from the drop-down list, select
the last item on the list and enter a custom answer in
the text box which appears.

Which country do you practice in?
New Zealand

Please indicate the highest educational
level in ANY area of study.
Bachelors Degree

Please select your status.
Practitioner (Registered/Licensed if applicable)

How many years of practice as an osteopath
have you had since graduation?
5

Please indicate how often you use visual
inspection of posture in your clinical reasoning
Always
Appendix 8. Example of video sequence of subject from CD-ROM.
Appendix 9. Freeze frames of video footage of subjects with varying degrees of forward head posture.

Appendix 10. Example of save and close option.
Appendix 11. Cohen’s kappa (linear weighted) inter-rater statistics for each individual observer.

<table>
<thead>
<tr>
<th>Laypeople</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
<th>Laypeople</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.184</td>
<td>0.173</td>
<td>9</td>
<td>0.054</td>
<td>0.165</td>
</tr>
<tr>
<td>2</td>
<td>-0.022</td>
<td>0.173</td>
<td>10</td>
<td>-0.137</td>
<td>0.165</td>
</tr>
<tr>
<td>3</td>
<td>-0.105</td>
<td>0.186</td>
<td>11</td>
<td>0.000</td>
<td>0.168</td>
</tr>
<tr>
<td>4</td>
<td>0.077</td>
<td>0.180</td>
<td>12</td>
<td>0.012</td>
<td>0.161</td>
</tr>
<tr>
<td>5</td>
<td>-0.182</td>
<td>0.187</td>
<td>13</td>
<td>-0.048</td>
<td>0.184</td>
</tr>
<tr>
<td>6</td>
<td>0.020</td>
<td>0.172</td>
<td>14</td>
<td>-0.168</td>
<td>0.215</td>
</tr>
<tr>
<td>7</td>
<td>-0.074</td>
<td>0.128</td>
<td>15</td>
<td>0.167</td>
<td>0.172</td>
</tr>
<tr>
<td>8</td>
<td>-0.063</td>
<td>0.138</td>
<td>16</td>
<td>-0.050</td>
<td>0.156</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
<th>Student</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.237</td>
<td>0.176</td>
<td>21</td>
<td>-0.206</td>
<td>0.179</td>
</tr>
<tr>
<td>2</td>
<td>0.050</td>
<td>0.142</td>
<td>22</td>
<td>0.029</td>
<td>0.153</td>
</tr>
<tr>
<td>3</td>
<td>0.210</td>
<td>0.168</td>
<td>23</td>
<td>0.381</td>
<td>0.170</td>
</tr>
<tr>
<td>4</td>
<td>0.138</td>
<td>0.172</td>
<td>24</td>
<td>0.140</td>
<td>0.195</td>
</tr>
<tr>
<td>5</td>
<td>0.058</td>
<td>0.174</td>
<td>25</td>
<td>-0.161</td>
<td>0.147</td>
</tr>
<tr>
<td>6</td>
<td>-0.077</td>
<td>0.171</td>
<td>26</td>
<td>0.039</td>
<td>0.176</td>
</tr>
<tr>
<td>7</td>
<td>-0.033</td>
<td>0.146</td>
<td>27</td>
<td>-0.273</td>
<td>0.160</td>
</tr>
<tr>
<td>8</td>
<td>-0.243</td>
<td>0.162</td>
<td>28</td>
<td>0.000</td>
<td>0.164</td>
</tr>
<tr>
<td>9</td>
<td>0.021</td>
<td>0.183</td>
<td>29</td>
<td>-0.191</td>
<td>0.167</td>
</tr>
<tr>
<td>10</td>
<td>0.142</td>
<td>0.115</td>
<td>30</td>
<td>0.029</td>
<td>0.125</td>
</tr>
<tr>
<td>11</td>
<td>-0.017</td>
<td>0.134</td>
<td>31</td>
<td>0.000</td>
<td>0.147</td>
</tr>
<tr>
<td>12</td>
<td>-0.138</td>
<td>0.162</td>
<td>32</td>
<td>0.075</td>
<td>0.168</td>
</tr>
<tr>
<td>13</td>
<td>0.017</td>
<td>0.167</td>
<td>33</td>
<td>0.028</td>
<td>0.151</td>
</tr>
<tr>
<td>14</td>
<td>-0.031</td>
<td>0.168</td>
<td>34</td>
<td>0.235</td>
<td>0.154</td>
</tr>
<tr>
<td>15</td>
<td>-0.061</td>
<td>0.172</td>
<td>35</td>
<td>0.146</td>
<td>0.162</td>
</tr>
<tr>
<td>16</td>
<td>0.169</td>
<td>0.154</td>
<td>36</td>
<td>-0.250</td>
<td>0.164</td>
</tr>
<tr>
<td>17</td>
<td>-0.092</td>
<td>0.171</td>
<td>37</td>
<td>0.101</td>
<td>0.152</td>
</tr>
<tr>
<td>18</td>
<td>0.123</td>
<td>0.171</td>
<td>38</td>
<td>-0.152</td>
<td>0.137</td>
</tr>
<tr>
<td>19</td>
<td>0.406</td>
<td>0.155</td>
<td>39</td>
<td>0.091</td>
<td>0.174</td>
</tr>
<tr>
<td>20</td>
<td>-0.111</td>
<td>0.154</td>
<td>40</td>
<td>0.087</td>
<td>0.172</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Practitioner</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
<th>Kappa (Linear weighting)</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.109</td>
<td>0.172</td>
<td>12</td>
<td>-0.139</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.170</td>
<td>13</td>
<td>-0.256</td>
</tr>
<tr>
<td>3</td>
<td>-0.151</td>
<td>0.168</td>
<td>14</td>
<td>-0.111</td>
</tr>
<tr>
<td>4</td>
<td>-0.323</td>
<td>0.165</td>
<td>15</td>
<td>-0.142</td>
</tr>
<tr>
<td>5</td>
<td>0.122</td>
<td>0.155</td>
<td>16</td>
<td>0.023</td>
</tr>
<tr>
<td>6</td>
<td>0.103</td>
<td>0.162</td>
<td>17</td>
<td>0.013</td>
</tr>
<tr>
<td>7</td>
<td>-0.092</td>
<td>0.157</td>
<td>18</td>
<td>-0.017</td>
</tr>
<tr>
<td>8</td>
<td>0.034</td>
<td>0.177</td>
<td>19</td>
<td>-0.045</td>
</tr>
<tr>
<td>9</td>
<td>0.114</td>
<td>0.165</td>
<td>20</td>
<td>0.236</td>
</tr>
<tr>
<td>10</td>
<td>0.250</td>
<td>0.162</td>
<td>21</td>
<td>-0.073</td>
</tr>
<tr>
<td>11</td>
<td>0.165</td>
<td>0.141</td>
<td>22</td>
<td>-0.184</td>
</tr>
</tbody>
</table>
Appendix 12. Response time versus subject number graph.
Appendix 13. The Instructions for Authors for the journal: Manual Therapy.

http://www.manualtherapyjournal.com/authorinfo

**Word Count**
Manuscripts should not exceed the following word counts:
- Original articles and review articles: 3500 words
- Technical and measurement notes: 2000 words
- Professional issues: 2000 words
- Masterclass: 4000 words
- Letters to the Editors: 500 words

These word counts do not include references or figures/tables.

**Presentation of Typescripts**
Your article should be typed on one side of the paper, double spaced with a margin of at least 3cm. One copy of your typescript and illustrations should be submitted and authors should retain a file copy. Rejected articles will not be returned to the author except on request.

Authors are encouraged to submit electronic artwork files. Please refer to http://www.elsevier.com/authors for guidelines on the preparation of electronic artwork files. To facilitate anonymity, the author’s names and any reference to their addresses should only appear on the title page. Please check your typescript carefully before you send it off, both for correct content and typographic errors. It is not possible to change the content of accepted typescripts during production.

Papers should be set out as follows, with each section beginning on a separate sheet: title page, abstract, text, acknowledgments, references, tables, and captions to illustrations.

**Title**
The title page should give the following information:

- * title of the article
- * full name of each author
- * you should give a maximum of four degrees/qualifications for each author and the current relevant appointment
- * name and address of the department or institution to which the work should be attributed
- * name, address, telephone and fax numbers, and e-mail address of the author responsible for correspondence and to whom requests for offprints should be sent.

**Keywords**
Include three or four keywords. The purpose of these is to increase the likely accessibility of your paper to potential readers searching the literature. Therefore, ensure keywords are descriptive of the study. Refer to a recognised thesaurus of keywords (e.g., CINAHL, MEDLINE) wherever possible.

**Abstracts**
This should consist of 150-200 words summarizing the content of the article.

**Text**
Headings should be appropriate to the nature of the paper. The use of headings enhances readability. Three categories of headings should be used:

- major ones should be typed in capital letter in the centre of the page and underlined
- secondary ones should be typed in lower case (with an initial capital letter) in the left hand margin and underlined
- minor ones typed in lower case and italicised

Do not use ‘he’, ‘his’ etc. where the sex of the person is unknown; say ‘the patient’ etc. Avoid inelegant alternatives such as ‘he/she’. Avoid sexist language.

**References**
The accuracy of references is the responsibility of the author. In the text your reference should state the author's surname and the year of publication (Smith 1989). If there are two authors you should give both surnames (Smith and Black 1989). When a source has more than two authors, give the name of the first author followed by ‘et al’. A list of all references in your manuscript should be typed in alphabetical order, double spaced on a separate sheet of paper. Each reference to a paper needs to include the author's surname and initials, full title of the paper, full name of the journal, year of publication, volume number and first and last page numbers.

**Citing and listing of Web references.**
As a minimum, the full URL should be given. Any further information, if known (Author names, dates, reference to a source publication, etc.), should also be given. The date on which the website was last accessed should also be included. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list. When citing a Churchill Livingstone journal, the digital object identifier (DOI) may also be included, if noted, from the article’s title page. Please note the following example: Joos U, Kleinheinz J 2000 Reconstruction of the severely resorbed (class VI) jaws: routing or
A detailed guide on electronic artwork is available on our website: http://www.elsevier.com/authors

Tables
Number tables consecutively in accordance with their appearance in the text. Place footnotes to tables below the table body and indicate them with superscript lowercase letters. Avoid vertical rules. Be sparing in the use of tables and ensure that the data presented in tables do not duplicate results described elsewhere in the article. Ensure that each table is cited in the text.

Preparation of supplementary data. Elsevier now accepts electronic supplementary material (e-components) to support and enhance your scientific research. Supplementary files offer the Author additional possibilities to publish supporting applications, movies, animation sequences, high-resolution images, background datasets, sound clips and more. Supplementary files supplied will be published online alongside the electronic version of your article in Elsevier Web products, including ScienceDirect: http://www.sciencedirect.com

In order to ensure that your submitted material is directly usable, please ensure that data is provided in one of our recommended file formats. Authors should submit the material in electronic format together with the article and supply a concise and descriptive caption for each file. For more detailed instructions please visit our artwork instruction pages at http://www.elsevier.com/authors

Submitting Case Reports
The purpose of the Case Report is to describe in reasonable detail the application of manual therapy to a clinical use. Cases of particular interest are those of an unusual presentation, rare conditions or unexpected responses to treatment. The following points will assist authors in submitting material for consideration by the Editorial Committee:

* The Case Report should be between 1500 - 2000 words in length excluding references and illustrations.
* Longer studies will be considered by the Editorial Committee if of an exceptional quality.
* The introductory paragraph should provide the reader with an overview of the study in general.
* The method of presentation to the treating practitioner should be detailed along with the symptoms and their behaviour. A body chart illustrating the symptoms is considered essential.
* The history (present and past) should be reported. Relevant work and leisure activities should also be presented in this section.
* The objective examination findings should be detailed in a concise manner.
* Treatment of the condition should be reported along with results. It is essential to clearly state what was done to achieve the reported results.
* The management of the condition should then be discussed with references to the literature to support what was done. Authors should remember it is a reasoned article rather than a purely factual report.
* The Case Report should conclude with a brief summary.

For further details on the Case Report section please contact: Jeffrey D. Boyling, Jeffrey Boyling Associates, Broadway Chambers, Hammersmith Broadway, LONDON, W6 7AF, UK. Tel: +44 (0) 20 8748 6878 Fax: +44 (0) 20 8748 4519 E-mail: jeffboyling@yahoo.co.uk

Submitting a Masterclass
The purpose of the Masterclass section is to describe in detail clinical aspects of manual therapy. This may relate to specific treatment techniques, a particular management approach or management of a specific clinical entity.

* The article should be between 3500 - 4000 words in length excluding references.
* A short summary should precede the main body of the article overviewing the contents.
* The introduction should review the relevant literature and put the subject matter into context.
* The main body of the text will describe the technique or approach in detail.
* Clinical indications and contraindications should be outlined when relevant.
* Illustrations are considered an essential part of the Masterclass in order to fully inform the reader and a minimum of six photographs or line drawings are required.
* Three copies of the Masterclass are required.

For further details and full instructions for authors for the Masterclass section please contact: Karen Beeton, Department of Physiotherapy, University of Hertfordshire, College Lane, HATFIELD, Herts, AL10 9AB, UK. Tel: +44 (0)1707 284114 Fax: +44 (0)1707 284977 E-mail: k.s.beeton@herts.ac.uk

Copyright Information
A "Transfer of Copyright" agreement will be sent to authors following acceptance of a paper for publication. A paper is accepted for publication on the understanding that it has not been submitted simultaneously to another journal in the English language. All authors must sign the "Transfer of Copyright" agreement before the article can be published. This transfer agreement enables Elsevier Science Ltd to protect the copyrighted material for the authors, without the author relinquishing his/her proprietary rights. The copyright transfer covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microfilm or any other reproductions of a similar nature, and translations. It also includes the right to adapt the article for use in conjunction with computer systems and programs, including reproduction or publication in machine-readable form and incorporation in retrieval systems. Authors are responsible for obtaining from the copyright holder permission to reproduce any material for which copyright already exists.
Funding body agreements and policies
Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier, to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies please visit http://www.elsevier.com/fundingbodies

Permissions Information
Written permission to produce borrowed materials (quotations in excess of 100 words, illustrations and tables) must be obtained from the original copyright holders and the author(s), and submitted with the manuscript. Borrowed materials should be acknowledged in the captions as follows: ‘Reproduced by kind permission of (publishers) from (reference)’.

Page Proofs
When your manuscript is received by the Publisher it is considered to be in its final form. Proofs are not to be regarded as “drafts”. One set of page proofs in PDF format will be sent by e-mail to the corresponding Author, to be checked for typesetting/editing. No changes in, or additions to, the accepted (and subsequently edited) manuscript will be allowed at this stage. Proofreading is solely your responsibility.

A form with queries from the copyeditor may accompany your proofs. Please answer all queries and make any corrections or additions required. The Publisher reserves the right to proceed with publication if corrections are not communicated. Return corrections within 48 hours of receipt of the proofs. Should there be no corrections, please confirm this.

Elsevier will do everything possible to get your article corrected and published as quickly and accurately as possible. In order to do this we need your help. When you receive the (PDF) proof of your article for correction, it is important to ensure that all of your corrections are sent back to us in one communication. Subsequent corrections will not be possible, so please ensure your first sending is complete. Note that this does not mean you have any less time to make your corrections, just that only one set of corrections will be accepted.

Author Enquiries
For enquiries relating to the submission of articles (including electronic submission where available) please visit http://www.elsevier.com/authors There is also the facility to track accepted articles and set up e-mail alerts to inform you of when an article’s status has changed, as well as detailed artwork guidelines, copyright information, frequently asked questions and more at: http://authors.elsevier.com/TrackPaper.html. Contact details for questions arising after acceptance of an article, especially those relating to proofs, are provided when an article is accepted for publication.

Checklist
Before submitting your paper, please check that:

* All files are uploaded.
* The reference list is complete and in correct style.
* Written permission from original publishers and authors to reproduce any borrowed material has been obtained.