Local Listening - a General Diagnostic Tool?
An Experimental Examination of its Reliability

Master Thesis zur Erlangung des Grades
Master of Science in Osteopathie

an der Donau Universität Krems,
niedergelegt an der
Wiener Schule für Osteopathie

von Wolfgang PODLESNIC

Wien, November 2006

Statistical Evaluation and Translation: Dr. Gebhard Woisetschläger

PREFACE

I would like to thank all those, who helped directly or indirectly in the finalisation of this master thesis, regardless whether this support was practical or mental.

In particular, I would like to thank Dr. Gebhard Woisetschläger for the statistical evaluation, many helpful suggestions and the translation of this thesis. His numerous valuable hints always contributed to an improvement of this work.

ABSTRACT

Agreements of findings in manual diagnosis often are low and thus reliability is poor.
The abdominal local listening test is one of the screening tests performed first by many osteopaths during the diagnostic process, and therefore in many cases an initial step for the further treatment. I want to find out via this methodological study, whether the same or at least a similar result can be obtained, when several osteopaths perform this test independently. It is my intention, to investigate the saying and reliability of this test generally and not only restricted to the visceral level.

14 osteopaths perform this test on 15 test persons, three of them twice. The direction where their sensing hand is drawn to is noted and the agreement of the sensed direction is evaluated by means of Cohen's kappa (κ-indices).
In approximately half of all comparisons between the therapists, the κ-indices indicate only agreements on the level of chance. The best agreements observed feature only a fair reliability.
The influence of the therapists' experience on the results did not turn out to be significant, though there might be trends that the agreement increases with the experience.
The number of agreements above the level of chance is higher in the first half of the investigation than in the second half. There is evidence that the conditions are changed during the examinations.
The values of the intra-examiner reliability are higher than those of the inter-examiner reliability (maximum moderate), but also here results can be observed, which are on a level of chance.
Summing it up, this test is not universally valid in the way as it was performed.
Additional supervision in advance of the examinations, a more distinct regulation of the level, from which the information should be gained, a longer time for the test and a regulation of the pressure of the hand should increase the agreements.

Table of Contents

1. Introduction and Goal of this Study

The abdominal local listening test is one of the screening tests performed first by many osteopaths during the diagnostic process, and therefore in many cases an initial step for the further treatment.

I want to find out via this methodological study, whether the same or at least a similar result can be obtained, when several osteopaths perform this test independently. By that I want to draw conclusions, whether this test is *universally valid* for osteopathic diagnosis, or only *subjectively valid* for the entire osteopath. It is my intention, to investigate the saying and reliability of this test generally and not only restricted to the visceral level.

Most osteopaths' experience corresponds with the following quotation of BECKER:

> *"It is interesting to note that the fact that I could not feel in the early years, did not determine the efficiency of the treatment."* (BECKER, 1997, 143)

In contrary to that is STILL's opinion, which seems to meet the specifications of modern science:

> *„He feels that the people expect more than guessing of an osteopath."* (STILL, 1986, 57)

Proceeding on the assumption, that the healing of the primary lesion is of main importance for an enduring solution of a health problem, methods to locate this primary lesion are of great importance.

Criticism of manual palpation techniques often is used to question their clinical relevance. A proof of good agreement, also under restricted but clearly defined conditions with regard to consistency, reproducibility and reliability, would support the osteopaths in their selection of therapeutic means. By that, an optimised therapy would serve the safety of the patients and be a basis for a successful treatment.

The abdominal local listening test was chosen due to its easy applicability and its possible key function for the further diagnostic process.

In the actual test, only the general direction where the therapist's hand is drawn (four quadrants) or a lack of a drawing shall be evaluated, respectively (Information about the test cf. chapter 2.2.1-2.2.3).

It is the goal of this study to find out, whether the local listening test is universally valid for osteopathic diagnosis or only relevant for a subjective diagnosis for each individual osteopath. It is my intention, to investigate the saying and reliability of this test generally and not only restricted to the visceral level. By this, a reliable decision about the selection of therapeutic possibilities shall be made feasible, serving the safety of patients and the success of therapy.

2. Basicals

Aside of an exact knowledge of anatomical and physiological fundamentals, sensing of deviations from them is a prerequisite for osteopathic treatment.
The information in the specific literature will be summarised in the following chapter, thematically subdivided into “Sensing and General Fundamentals of Tests”, “Local Listening” and “Manual Diagnostics”.

2.1. Sensing and General Fundamentals of Tests

2.1.1. Sensing from a Neurophysiological Point of View

Manual diagnosis is one of the most important instruments for osteopaths. Thus the objectivity of the sensed is of great importance.

The human hand is endowed with certain neurons that are primarily sensitive to positional changes and other neurons sensitive to the time rate of positional change or velocity (KNEIBESTOL AND VALLBO, 1970). Thus the sensation appreciated by the palpating hand could be derived from a combination of both positional and velocity factor (also compare BURGESS AND PERL, 1973 and VALLBO AND HAGARTH, 1968).

WEBER (2000) compares the physiological influences of body position, breathing, thoughts and therapist’s intention on sensing with osteopathic philosophy and concludes that for accurate sensing the osteopath has to be unrestrained by stress, fear, patterns of flight, incorrect concepts of thinking and egoistic consciousness.

CLOET (2005) also refers to the last point in a contra-statement in a discussion about the importance of visual perception:

> *„ [...] because the knowledge about the patient, which we derive by palpation, is spinning a guiding thread in our brain, which is referring to our own knowledge. We adjust our behaviour according to this guideline, because our ego demands this“* (CLOET, 2005, 29 translated by WOISETSCHLÄGER, 2006).

Further he concludes:

> *„If we want to work rationally, it is better not to use the different sense organs simultaneously as often as possible. Brain research has proven that „multitasking“ is bringing about stress and this is hardly worthwhile.“*
>
> *„Tasks, we perform automatically, require less mental aid but can not be performed without a certain attentiveness. And to switch from one task to another has its price: one can focus less onto that, one is occupied with and loses time during „resetting“* (CLOET, 2005, 29 translated by WOISETSCHLÄGER, 2006).

2.1.2. Requirements for Tests from a Scientific Point of View

For a scientific approach, reproducibility, consistency and reliability are of great importance for all possibilities for gaining information in the diagnostic field (instrumental diagnosis, anamnesis, ...). Naturally, this also applies to manual diagnostic techniques. A decisive factor for the course of therapy for a patient should fulfil all these criteria.
So, what do these criteria mean?

Reproducibility means that a measurement can be repeated under the same conditions. This reproducibility is required from one and the same person or one and the same measuring instrument (intra-examiner reproducibility), as well as from different persons or measuring instruments *(interexaminer reproducibility*) (SOMMERFELD, 2000, 4).
Totally equal conditions are only difficult to ensure in the dynamic systems of the patients as well as the therapists.
In this case, when the results cannot be repeated as often or exact as wanted due to influences, which are not checked or difficult to check, statistical methods have to be used (SOMMERFELD, 2000, 4).

Influences on the patients by postulated therapy effects (RUSSEL, 1983 and GOLDSTEIN, 1978) by palpation or on the diagnostician by other information flows de facto cannot be precluded. Therefore, a priori, lower reproducibilities are adjudged to manual methods than to instrumental ones. Nevertheless, some studies in the literature show, that seemingly “subjective” methods are more reliable than “objective”

ones, among them also instrumental (CF. KEATING ET AL., 1990; BOLINE ET AL., 1993; BOLINE ET AL., 1988 and MOOTZ ET AL., 1989).

In order to guarantee, that as little information as possible can be exchanged between diagnosticians, studies have to be blinded.
Therapy effects by touching the patient during diagnosis can only be precluded by simultaneous diagnosis by all therapists, which hinders the blinding.

The reproducibility of the results of one and the same therapist can only be evaluated by means of sequentially performed palpations. Thus, the prerequisite of the same conditions might not be fulfilled for the patient as well as the therapist.
Additionally, HAAS (1991) doubts the possibility of blinding of one and the same examiner against the same patient (HAAS, 199-208).
Due to reasons of therapy effects as well as self-healing processes the reproducibility should increase with the seriousness of the symptoms.

In the actual study, intra-examiner reproducibility will be investigated by means of the agreement of the three osteopaths who do the first three tests on the test persons and repeat the test when all other osteopaths have finished their tests.
The inter-examiner reproducibility will be calculated for pairs of different osteopaths. The study will be performed blinded. Nevertheless, therapy effects cannot be precluded, since the tests will be performed sequentially.

Consistency is a measure for the exemption from contradictions which can be tested, if different contradictory statements occur, which are equally registered by different individual observers (SOMMERFELD, 2000, 5).
In the actual study consistence is reached, when each osteopath feels the same quadrant at one and the same test person, and another quadrant at another test person without contradictions of the individual therapists.

Reliability is a measure which covers as well reproducibility as also consistence. It defines the efficiency more accurately by defining criteria for the circumstances of the measurements. Also here, analogous to reproducibility it is distinguished between *Inter- and intra-examiner reliability* (SOMMERFELD, 2000, 5-6).

The reliability can be evaluated by testing the measurement outcomes under these conditions and by optimising the conditions it can be improved, respectively.

Summing up, it may be said that the definition of conditions and an exactly defined course of a study are absolutely necessary for a study about a manual diagnostic technique.
Further, conditions have to be prepared, in which osteopaths can work unrestrained by stress, fear, patterns of flight, incorrect concepts of thinking and egoistic consciousness.

2.2. The Local Listening Test

2.2.1. Fundamentals of the Local Listening

BECKER (1997) was the first who used the term “listening to the tissue”, which was described by BARRAL as “Ecouté” (fr. listening).
Shortly said, according to BARRAL (BARRAL AND MERCIER, 1988), who developed the entire concept of listening, it is the sensing of global or local tissue tensions by means of the sensitivity of the hand.

BARRAL (BARRAL AND MERCIER, 1988) assumes - compatible with the osteopathic principles - (cf. Chapter 2.2.2.3.), that each organ or viscera in a good state of health has a physiologic motion, which is an interdependent one due to the serous membranes which envelop the organ and the fasciae, ligaments and other living tissues which bind it to the rest of the organism.

The physiological movement can be divided into two components: (1) ***visceral mobility*** (movements of the viscera in response to voluntary movement, or to movement of the diaphragm in respiration) and (2) ***visceral motility*** (inherent motion of the viscera themselves).

According to BARRAL in „Visceral Osteopathy“ global listening is described as *“’essential modality’* for the evaluation of the axis and amplitude of motility of any viscera.” (BARRAL AND MERCIER, 1988, 21).

In „Visceral Osteopathy II“ (BARRAL, 2002, 5-9) two further tests are introduced by BARRAL, also under the same term “ecouté”, which can lead to mistakes. These tests are the “global listening” and “local listening”.

In the actual study the **technique of the local listening test** (BARRAL, 2002) is used, which allows to localise global or local tissue tensions. By means of this test as well mobility as also motility of the viscera are considered, in order to find reference points for functional impairments and consequently for possible therapies.
Due to the bounds of the body structures not only visceral impairments, but also those in other body structures can be localised.

Ecouté means listening to the tissue, covering different methods. The actual study bases on the technique of local listening according to BARRAL (2002), a method, which enables to perceive the condition of the body or a certain organ or viscera.

2.2.2. Visceral Mobility and Motility

The basis of the local listening test is the sensing of body or tissue tensions, which indicate changes in visceral mobility and motility (and affect them at the same time). These kinds of movement and their restrictions will be summarised in the following sections.

2.2.2.1 Visceral Mobility

According to BARRAL (BARRAL AND MERCIER, 1988), every restriction of the mobility affects the structures and organs which are affected by it. This fact is compatible with the osteopathic principles (cf. chapter 2.2.2.3).

Voluntary muscular movements controlled by the central nervous system lead to a passive movement of the viscera due to the various anatomical articulations. The nature of these articulations determines the axes and the amplitudes of voluntary movements. Due to the anatomical fundaments these movements can be predicted and deviations can be sensed.
Also autonomous movements, as for example the piston-like movement of the diaphragm or the pumping of the heart, lead to passive movements of the related body structures

These autonomous movements, repeated thousands of times daily in the body, bring about sliding and rubbing phenomena of the viscera in the three cavities (pleural, pericardial and peritoneal) of the trunk.
Peristaltic motion is another autonomous movement and brings about a stirring and circulating the visceral contents with its great contractile waves.
BARRAL also mentions the craniosacral rhythm, which also affects the bones of the skull, as well as other skeletal structures (BARRAL AND MERCIER, 1988).

2.2.2.2 Visceral Motility

Aside of the passive motions influenced by extrinsic factors the viscera also have an intrinsic, active motion, according to BARRAL (BARRAL AND MERCIER, 1988), which is called motility.
They move independently, with a motion which is slow and of such low amplitude as to be almost imperceptible and according to this model are a measure for the vitality of the organ.

Visceral motility is perceptible to the hand but requires an educated sense of touch.
BARRAL points out that there is no scientific explanation for this phenomenon, and speculates, whether it is an extension of the craniosacral rhythm, or it corresponds to movements of the organs during embryogeny (BARRAL AND MERCIER, 1988).

The embryologic theory of visceral motility postulates that the axes and directions of these motions remain inscribed in the visceral tissues. Thus, visceral motility occurs around a point of equilibrium, oscillating between an accentuation of the embryologic motion and are turn to the original position, with a contractility analogous to (but much slower than) that of the nodal tissue of the heart.
The motility cycle has two phases, in which the organs move toward and away from the median axis of the body. These phases are called "expir" and "inspir" respectively. Under normal conditions the organs move in sync, i.e., they all undergo inspir or expir at the same time. There is no particular relationship between the direction of motion of the organs during the different phases of visceral mobility and those in the phases of visceral motility (BARRAL AND MERCIER, 1988).

The cranial theory is built on the similarity of the craniosacral flexion and extension with inspir and exspir of motility, as well as the activities during these phases.

Nevertheless, BARRALL does not describe this theory concisely (BARRAL AND MERCIER, 1988).

2.2.2.3 Restrictions of Mobility and Motility and their Consequences

Mobility and motility in equilibrium of direction and amplitude can be rated as a sign for a dynamic balance in the body.

BARRAL (BARRAL AND MERCIER, 1988) distinguishes between „functional restrictions“ and „positional restrictions“ in consequence of deviations from this equilibrium, and between „articular“, „ligamentous“ and „muscular“ restrictions in consequence of the mode of the restriction.
With **functional restrictions***,* only the function of the related organs is affected; their positional relationships are not changed.
With **positional restrictions***,* the anatomical relationships of the organs are changed and their articulations are modified, in rare cases it is also possible to have a positional restriction without a functional restriction.

Articular restrictions are restrictions, which are designated by Barall as **„fixation“** or **„adhesions“**. The first describes a restriction of mobility and motility, the latter only a reduced motility.
The reasons for them are natural or surgical healing processes involving the local disruption of normal tissue fibers and their replacement with relatively inelastic granular tissue. Partial fixations will only modify the axis of motility. The axis goes through the adhesion. Total articular restrictions completely inhibit motility; the organ becomes inert and loses its rhythm. In consequence, its vitality and ability to function properly are affected.

Scars (as a consequence of infections or operations) create a permanent state of mechanical irritation by forcing the tissues to rub against each other resulting in a pathological decrease in motion over time.
Attached tissues and organs will find their axes modified, the circulation of blood and lymph through the organs will be diminished, and consequently the immune system is affected.

Ligamentous restrictions are a loss of elasticity in the ligament from prolonged overstretching, usually secondary to adhesions.

Beside others, some reasons given by Barral (BARRAL AND MERCIER, 1988) are hypotonia, underweight, a decrease in tonus due to depression, aging and multiparity.

Muscular Restrictions almost exclusively affect the hollow organs, which consist of a double smooth musculature with longitudinal and transverse circular fibres. Irritation of a group of fibres can make them go into spasm.
With muscular restrictions, motility is affected first; mobility is only decreased when the organ's attachments are affected.
The effects of these restrictions can be observed in alterations of the axes of motion and subsequent rubbing tissues, disruption of tissue fibres and positional changes (and finally changes of mobility) of attached organs.

BARRAL AND MERCIER (1988) write:

> *„A small disturbance in motion, repeated millions of times over months or years, can provoke problems seemingly disproportionate to the original cause."* (BARRAL AND MERCIER, 1988, 21).

Motility is changed by adhesions and restrictions in the surrounding tissues which change the axes, upset the symmetry, and decrease the amplitude of motility. If the amplitude of motility decreases drastically it often becomes fixed in expir.

Due to the structural attachments in the body and the mutual influence of structure and function, these effects are not confined to visceral structures, but influence also all other body structures.
This shall be displayed by the osteopathic principles by Still (1986) (quoted by RUSPECKHOFER, 2000, 33-35, original literature cf. STILL, 1986):

> ***Principle 1: Life is motion***
> *Motion is one of life's basic principles - in a mechanical sense, as well as in the senses of dynamics, alterations, action and behaviour. Motion can be seen as a means of measuring the vitality of the whole organism, as well as of its parts, down to cellular and molecular level.*

Principle 2: Structure and function are in a reciprocal relationship

The mutual influence of structure and function is found on:

- *The mechanical level, between joints, muscles and bones.*
- *The membranous level through fascial and ligament connections between organs and tissue.*
- *The circulatory level due to the course of blood- and lymphatic vessels and the fluctuations of the fluids of the cerebrum and the spinal cord.*
- *The neurological level through information transfer of the peripheral- and central nervous tract.*
- *The biochemical, hormonal and electro-physiological level between tissue and organs.*

A regular structure and a physiological tension of all of the body's tissue are necessary to assure optimal function.

The vascular systems and nerves provide an integrated and supporting framework for the whole organism.

Principle 3: The body functions as an entity

All the cells, tissue and organs of the body co-operate. Abnormal structural changes or disturbances in function of a single tissue might affect the whole organism.

Principle 4: The law of the arteries

A proper circulation of all bodily fluids (blood, lymph, liquor, and gas exchange) is a prerequisite for health. The loss of mechanical flexibility and hyper-tension inside the tissue lead to a diminution of the dynamic behaviour of body fluids, and finally to a deterioration of the supply- and discharge condition. Problems tend to develop in less supported tissues. Apart from the defective support, the disturbed removal of toxic substances accounts for this situation.

Principle 5: Self healing mechanism

Good health is not a coincidence. It is the result of various autoregulation processes of the immune system, the endocrine system, the autonomous nervous system and other regulative systems. Through the therapeutic resolving of various pathogenic influences, the organism can gain enough compensatory possibilities and becomes able to heal itself. (STILL (1986), quoted by RUSPECKHOFER, 2000, 33-35)

Summing up, it can be said that adhesions and fixations restrict mobility as well as motility in their natural axis and amplitude. As consequence not only the vitality of the affected organs is reduced (“life is motion”), but also the structure and the state of tension will be affected by the mutual influence of structure and function.
The ability to function properly is not granted anymore. Since the body functions as an entity, these changes have effects on other structures and functions, too. The loss of mechanical flexibility and hyper-tension inside the tissue lead to a diminution of the dynamic behaviour of body fluids and as a consequence to a deterioration of the supply- and discharge condition.

Visceral mobility is the passive motion of the viscera, provoked by other movements (conscious but also autonomous ones, as breath and pumping of the heart).
Visceral motility is the inherent motion of the viscera themselves.
Visceral Mobility, as well as motility are dynamic processes, which can be restricted by adhesions and fixations in their natural axis and amplitude. These restrictions can lead to chains of lesions and disease, if they are not treated.

2.2.3. The Abdominal Local Listening Test: Realisation and Information Gained

The procedure of the local listening is described in „Visceral Manipulations II“ (BARRAL, 2002, 7) as follows:

> *“Work with your more sensible hand and in supine position of the patient. The hand should be placed in the median line on the abdomen, the, the ball of the thumb directly on the navel and the fingertips below the Proc. xiphoideus. Righthanders usually use their right hand and thus stand or sit on the right side of the patient. Your hand receives passively the information from the tissues. If a tissue is too tense, your hand will be attracted by it.*
>
> *You can feel, how the hand gradually moves towards the lesion. Sometimes it moves step-by-step. E.g. you can feel, that the lower border of your hand moves from the median line to the right border of the rip cave. Follow the movement […]. Nevertheless, if there is no major lesion, your hand will not be attracted by a certain point. Also at the final control after a treatment no attraction should be sensed anymore [...]* (BARRAL, 2002, 7, translated by WOISETSCHLÄGER, 2006).

Since this text is taken from a textbook for visceral techniques, I want to stress, that this test is not only suitable for viscera. In my opinion, also lesions in other structures, from which tissue tensions can arise, can be diagnosed.
It is the aim of this test, to find a first trace of lesions, by sensing movements towards irregularities in the tissue tension, which can be expressions of an impairment.
Disturbances of mobility, as well as motility can be realised by this test and - because body structures are inseparable - also influences of other non-visceral ones.

By the passive receiving of information from the tissues by the hand the sensing hand will be attracted by more tense tissue and it will move towards the lesion. From this information a first trace for the further diagnosis and subsequent therapy can be found.

2.3. Literature about Manual Diagnosis

The discussion about the reliability of manual diagnosis in osteopathy has begun just in the last years. Thus, only view studies have been published yet.

In the literature (CF. DINNAR ET AL., 1980; DINNAR ET AL., 1982 and KUCHERA AND KAPPLER, 2002) it is distinguished between four different fundamental types of palpatory tests: Differentiation of tissue textures, evaluation of static landmark positional asymmetry, evaluation of motion asymmetry and assessment of tenderness.

2.3.1. General Literature

In this chapter, fundamental aspects, which are of importance for the study design, are summarised.
In a review about reliability studies, ALLEY (1983) calls the attention to the fact, that studies about the test-retest reliability previous to 1983 suffer statistical and methodological deficiencies. Especially, he criticises the use of percent agreement without the consideration of agreement expected on the basis of chance (ALLEY, 1983, 97-100). The same deficits are criticised in RUSSEL (1983) and HAAS (1991).
In the introduction of their osteopathic study about neuromusculosceletal investigations, McCONNELL ET AL. (1980) refers to the result of KORAN, where the

reason for a lack of agreement between examiners was fixed on not previously agreed terminology, observational criteria and data collection protocols (KORAN, 1975).

MIOR ET AL. (1990) state a dependency of intra-examiner reliability, but not of interexaminer reliability of the experience of the examiner for the motion palpation of the iliosacral joint (MIOR ET AL., 8-71).

In contrary, according to BEAL AND PATRIQUIN (1995) also interexaminer reliability in osteopathic tests increases with the practical experience.

MUZZI (2005) summarizes the prerequisites for a good agreement of results of manual diagnosis in three points (MUZZI, 2005, 16, translated by WOISETSCHLÄGER, 2006):

- *Take only one and the same examiner.*
- *The examiner should have very good skills or start a pilot study and/or a consensus training, to refine his palpatory skills for specific tests.*
- *Make use of a blinded study for the analysis of the dependent variable* (MUZZI, 2005, 16).

Considering these inputs from the literature, I drew the following conclusions for the actual study:

- The study will be performed blinded.
- It will be stressed that the examiners must not treat at all.
- Data about the examiners' experiences will be collected.
- Measures will be taken in order to recognise treatment effects.
- The osteopaths' attention will be called to the fact, that not only visceral concerns are to be considered in advance of the test.
- In the evaluation, according to the present medical literature, Cohens' κ-indices will be used.

2.3.2. Literature about the Inter- and/or Intraexaminer–Reliability of Manual Tests

The literature research was performed at the central library of the medical faculty of the Vienna University, by searching the Online-catalogue of the medical periodicals, as well as medical online data banks for authors of already known articles and the following key words: Osteopath(y, ic), reliab(ility), reproducib(ility), motil(ity), mobil(ity test), ecout(é), listening, manual test, induction, chiropr(actic), craniosacral, tension.
Also parts of these key words and combinations of them were used.

In contrary to other manual diagnostic tests (e.g. palpation for mobility) no literature concerning „listening“ or „Ecouté” - neither of general interest nor about their reliability - can be found.
In order to be able to compare the results of this study, literature about manual diagnosis on the musculo-skeletal structure, about pulse palpations and about investigations of the cranio-sacral rhythm will be summarised on the following pages.

2.3.2.1 Literature about Manual Diagnosis on the Musculo-Skeletal Structure

In this connection, predominately literature about passive but also active mobility tests can be found. Anticipating the results, agreements of findings in manual diagnosis often are low and thus reliability is poor.
Only few passive tests, where also the „local listening“ can be counted to, are published in the literature. The following summaries of the literature results are sorted by the order of their publication.

McCONNELL AND COAUTHORS (1980) describe the results of their findings of neuromusculosceletal examinations on patients with acute pain in the spine. In agreement with the praxis-situation the examination techniques were not standardised, resulting in an individual choice of techniques.
Six graduated osteopaths investigate all segments from C0 to S1 of the spines of 21 patients and record their findings numerically on a scale of 0 to 3 with respect to clinical significance. Additionally, the therapists are asked about their working diagnosis, where they would initiate treatment, whether they would rate their finding acute or chronic and in what position the patient is during the examination.
The first osteopath takes the patient’s history, the following does not have access to these data.

Various statistical methods are used to assess the level of agreement among the examiners: First, expected disagreements on a single patient based on the uniform distribution hypothesis are calculated and secondly, agreement in cluster areas is computed. In the first case all possible differences between two physicians are compared with the actual differences and thus findings on the basis of change can be eliminated. In the second case the relative importance assigned to different regions is compared.
The result is a low agreement between the examiners, which additionally is dependent on the region and the acute segments.
Differences in examination technique, in the perception of importance of findings and in conceptual orientation are named as reasons for the unsatisfying results (McCONNELL ET AL., 441-450).

MYERS ET AL (1987) investigate the inter-examiner agreement of the palpation of the femoral and popliteal pulse. Six vascular surgeons examine 44 legs in patients with suspected peripheral arterial disease, whether pulses are present and whether they are normal or reduced in amplitude. Agreements as to whether pulses are present or absent are moderately good (κ = 0.53 for femoral pulse and κ = 0.52 for popliteal pulse). More often than not, agreement as to whether the pulses are normal or reduced is no better than expected by chance (κ = 0.15 for femoral pulse and κ = 0.01 for popliteal pulse) (MYERS ET AL, 245-249).

BOLINE ET AL (1988) investigate the interexaminer reliability of palpations in the lumbar region. 23 patients with low back pain and 27 asymptomatic test persons are assessed by two examiners using passive motion palpation for fixations and using static palpation for pain and muscle tension.
The agreements differ depending on the position, with lowest agreements in the lower lumbar spine. Reliability is higher at symptomatic patients than on asymptomatic test persons (BOLINE ET AL., 5-11).

MOOTZ ET AL (1989) investigate the intra- and inter-observer reliability of passive motion palpation of the lumbar spine.
Mobility of the segments L1/2, L2/3, L3/4, L4/5 and L5/S of 10 test persons is examined by two chiropractors in six directions.
The two by two agreement matrices are constructed for each motion segment and subsequently the κ-index is applied to each matrix.

Each examiner demonstrates significant test-retest agreement for palpatory evaluations at L1/2, (κ=0.39 and κ=0.48) and L4/L5 (κ=0.21, κ=0.29), interexaminer reliability is not significant at all. As already discussed at McCONNEL, agreement varies with regard to certain regions.
MOOTZ concludes that a documentation of existing techniques, a refinement and standardisation of more reliable and valid palpatory indicators of segmental dysfunctions are important in order to get more reliable results (MOOTZ ET AL., 440-445).

KEATING ET AL (1990) investigate the reliability of eight diagnostic dimensions per lumbar segment from Th11 to S1. Three examiners assess 25 asymptomatic and 21 symptomatic test persons by means of visual diagnosis, dermatograph, palpation of pain on the osseous and soft tissue, palpation of muscle tensions, static palpation and active and passive motion palpation.
Reliability is calculated by Kappa–indices (κ), intraclass correlation coefficients (ICC) and Pearson's correlation coefficient (r).
The results reveal that methods which are commonly judged as the most "subjective" ones, such as osseous and soft tissue pain palpation, are the most reliable (κ= 0.10 - 0.66), followed by dermatograph and visual observation. Static palpation, tests for muscle tension and active as well as passive motion palpation result in low reliability or even agreement only on the basis of chance (κ=-0.28 - 0.32) (KEATING ET AL, 463-470).

BOLINE ET AL. (1993) investigate the agreement of examiners on 28 patients with low back pain by means of visual, palpatory und instrumental diagnostic procedures (portable EMG – surface scanner, dermothermograph) for the lumbar spine. The data are evaluated by kappa – indices (κ) and percent agreement. In contrary to some manual diagnostic methods (good to excellent agreement for visual observation and palpation for pain), instrumental methods reach only missing to poor agreement (BOLINE ET AL, 363-374).

HUBKA AND PHELAN (1994) describe the interexaminer reliability of manual palpation of cervical spine tenderness in patients with unilateral neck pain. For this purpose, the cervical spines from C2 to C7 of 30 patients are palpated by two chiropractioners with well defined methods. Only the randomly chosen first therapist has access to anamnesis data. The repetition of the palpation by the second therapist

shows a substantial reliability with κ = 0.68 and p<0.001. These findings about the palpation of cervical spine tenderness thus are consistent with those of palpation of the lumbar spine (HUBKA. AND PHELAN, 591 – 595). (CF. KEATING ET AL, 1990, 463-470; BOLINE ET AL, 1993, 363-74, PANZER, 1992, 518-524 AND CARMICHAEL, 1987, 164-171).

LUNDIN ET AL. (1999) investigate the reliability of distal pulse palpation (A. dorsalis pedis and A. tibialis posterior). Nine examiners palpate the pulse of 25 patients with suspected lower limb arterial disease. The palpation findings are compared to the ankle/brachial index (ABI). The palpation technique is not standardised. The investigations are performed under two different conditions, an undisturbed examination situation with sufficient time and a hectic outpatient clinic.
The agreement of the palpation results with the ABI is estimated by percent agreement and kappa statistics. Different ankle/brachial indices are used as reference points for the separation of palpable from non-palpable arteries.
The proportion of underdiagnosis is 33% with an ABI of 0.96 as criterion of disease. Overdiagnosis occurs at 19%. With an ABI of 0.71 the results are 19% for underdiagnosis and 34% for overdiagnosis, respectively. With a mixed concept (ABI>0.96, underdiagnosis; ABI<0.71, overdiagnosis), the overall proportion of misdiagnosis is 18.8%. Individually, the examiners have percentages of severe misdiagnosis between 9.7% and 32.3%.
An undisturbed examination situation with sufficient time to examine each patient proves to be of primary importance. Under quiet conditions, the examiners reach a kappa of κ=0.68, whereas the agreement is lower with κ= 0.38 in the busy outpatient clinic (LUNDIN ET AL., 252-255).

HAWK (1999) performs a preliminary study, assessing the intra- und interexaminer reliability for the indication of chiropractic manual therapy for the segments of the lumbar spine (Th12/L1 to L5/S1).
Four chiropractors trained in flexion-distraction technique with different experience palpate 18 test persons by means of static and motion palpation and visual observation. In order to reproduce the „real-life“ clinical setting, each examiner are allowed to use his own combination of techniques..
The κ-indices are calculated for all comparisons. The intra-examiner reliability is higher than inter-examiner reliability, which generally is in the “poor” to “slight” categories (HAWK ET AL., 382-389).

SCHÖPS ET AL (2000) investigate the reliability of manual examination at the cervical spine. 20 patients suffering from neck diseases and 20 asymptomatic test persons are randomised and assessed by five examiners blind to patient histories. The cervical zygapophysial joints and the superficial neck muscles are tested for pressure sensitivity and a segmental function test of the segments C0/C1 to C7/Th1 is performed regarding to hypomobility and pain.
Significant findings are described only for the palpation of tension in joint-facets and superficial neck muscles as well as for induced kinesalgia. No significant relationship can be found between the patient's health status and the findings from muscle palpation and functional examination of the motion segments.
The reliability between the examiners turn out to be fair to moderate. ($0.2<\kappa<0.6$) (SCHÖPS ET AL., 2-7).

LAWSON AND CALDERON (1997) describe the inter-examiner reliability for applied kinesiology manual muscle tests. Three examiners with greater than 10 years experience test 32 healthy test persons to estimate their agreement on the strength or weakness of left and right m. piriformis and left and right hamstring muscle and 53 test persons for the strength or weakness of the m. pectorialis and the tensor fascia lata. The test person's laying position is exactly defined.
Again, the results vary structure-specific. The concordance among all examiner-pairs is good to excellent and statistically significant for the tests of the m. piriformii. The findings for the hamstring muscles are not statistically significant in three of five tests. Mixed results are achieved in the reliability trials of the tensor fascia lata. Also for the m. pectoralis a significantly good reliability is found. The different results for the test for hamstring muscles are explained by the fact that this is a group test rather than one of an individual muscle. Patterns in which one division of muscles is "weak" may lead to variable recruiting patterns yielding variable results (LAWSON AND CALDERON, 539-546).

HESTBAEK AND LEBOEUF-YDE (2000) review studies about the reliability and validity of chiropractic tests used to determine the need for spinal manipulative therapy of the lumbo-pelvic spine which were performed between 1976 and 1995, taking into account the quality of the studies.
In summary, only studies focusing on palpation for pain (palpation for tenderness) have consistently acceptable reliability values. These results are patient-induced, whereas clinician-induced interpretations are worse.

Palpatory tests of landmark positional asymmetry, movement asymmetries und muscle tension result in poor reliability.
Concerning the palpation for muscle tensions only BOLINE (1988) and KEATING (1990) meet the quality criteria of the authors (HESTBAEK AND LEBOEUF-YDE, 258-275).

DEGENHARDT ET AL (2005): The authors investigate the inter-examiner reliability of common osteopathic palpatory tests used to evaluate the lumbar spine in a double blinded study. 119 test persons are examined by three experienced osteopathic medical examiners in two subgroups. The osteopaths perform palpatory tests of tenderness and tissue texture changes, as well as - in three planes - vertebral positional asymmetry and motion asymmetry. In the first subgroup (n=42) κ-indices ranging from κ=–0.02 to 0.34 (within the poor to fair reliability range) are reached. After a following consensus training reliability increases in the second subgroup (n=77), rising to κ=0.45 for tissue texture changes (moderate), to κ=0.68 for tenderness (substantial). The reliabilities for positional asymmetries in the transverse plane and for rotational motion asymmetries improve only slightly (κ= 0.34, κ=0.20, respectively).
The authors conclude from the results of the consensus training, that osteopathic medical educators need to modify their curricula to better calibrate and standardize palpatory diagnostic skills (DEGENHARDT ET AL., 465-473).

HUDSWELL ET AL. (2005) describe connections between the results of the cranio-cervical flexion test (CCFT) with pain intensity, history of neck pain and the level of neck disability. The intra- and inter-examiner reliability of four therapists, who perform a CCFT on 40 test persons, demonstrate moderate inter-rater agreement and substantial intra-rater agreement for the performance index and the activation score. Nevertheless, the CCFT fails to discriminate between those with current neck pain, those with a history of neck pain but no current pain, and those without neck pain (HUDSWELL ET AL., 98-105).

The literature summarised above shows that manual diagnosis methods cover a high variability of reliability.
Dependencies of the amount of reliability on the body structure of the patient, as well as on the therapist's experience and ability to concentrate can be observed.
Some authors also mention the need of a better training during education.

In my opinion, it has to be considered, that in many cases therapy is successful even in spite of different results of osteopathic diagnosis, indicating that different therapy concepts may lead to success.

2.3.2.2 Literature about Manual Diagnosis of Rhythms

SMITH AND CRAIGE (1980) investigate the frequency response of the hand, to improve the sensitivity of the hand as transducer for precordial movement. The threshold of tactile sensation is determined for 10 test persons by variations of the amplitude of movement of an impulse generator (palpation simulator) at each of a series of frequency settings in the subaudible range (1-40 Hz). Since the ratio of maximum velocity to amplitude is proportional to frequency and the ratio of maximum acceleration to amplitude is proportional to the square of the frequency, by adjusting the frequency setting the relative contributions of the amplitude, velocity and acceleration can be altered.

The test persons rest a hand lightly on the disc of a palpation simulator, in a second experiment finger movements are restrained. The experiment with the restrained hand is a simulation of specific bimanual palpation techniques where the contacting hand stays relaxed whilst guidance and pressure is exerted by the second hand which is put on top of the first.

The subject increases gradually the amplitude setting of the signal until the threshold of sensation is obtained. The range of amplitudes is from several hundredths of a millimeter to approximately 1 mm, the frequencies cover the range with settings between 1 - 40 Hz.

The threshold amplitude decreases with increasing frequency (above 5 Hz), because owing to inertia, the free fingers are unable to follow the motion of the palpation disc as the frequency of the impulse increases.

With the restrained hand the same effect can be observed. But in this case, amplitude sensitivity is approximately 4.7 times higher than when the hand is allowed to float freely. This effect is highest in the lower frequency range (<5 Hz). The authors refer to BURGESS UND PERL (1973), who proved, that there are certain mechanoreceptors in mammalian skin, primarily sensitive to velocity, while others are sensitive to positional change. The mechanoreceptors also show a sensitivity to rapid transients of motion that correspond to high-frequency stimulation, suggesting that acceleration might be contributing factors to the total neuronal response of the tactile sense (BURGESS AND PERL, 30).

In the restrained hand also the sensory neurons on the back hand are stimulated additional to those of the volar surface of the fingers and thus the total neuronal response to the wave form is increased.
Additional pressure also leads to a further indentation of the skin, which is necessary for the detection of motion. Nevertheless, further experiments state that no comparable improvement can be obtained by attempting to restrain one hand with the other or with the application of heavy weights, because by the compression of the tactile sensory surface results sensory fatigue and a loss of perception. Also with the palpation simulator sensory fatigue begins to become a factor after approximately 30 seconds (SMITH AND CRAIGE, 1114-1118).

LOCKWOOD AND DEGENHARDT (1998) investigate the cycle-to cycle variability of the primary respiratory mechanism. They describe, that two physicians focus on the duration of each cycle but do not present their results unless the postulated consistency of their measurements and interexaminer agreement.
The duration of individual cycles demonstrate significant cycle to cycle variability of 0.6 seconds up to 6.3 seconds. The minute rate of each tracing ranged from 0.108 up to 0.230 Hz. The authors suggest that this variability has confounded previous interexaminer reliability studies (LOCKWOOD AND DEGENHARDT, 35-36, 41-43).

In a study by MORAN UND GIBBONS (2001) two osteopaths simultaneously investigate the phase of full flexion of the cranial rhythm impulse (CRI) on cranium and sacrum of 11 asymptomatic test persons. Additionally, the test person's pulse rates are registered. The agreement of the two osteopaths is in the fair to good reliability range (ICC of 0.52 to 0.73) at the same body region (cranium or sacrum), but only poor, when cranium and sacrum are taken into consideration simultaneously ICC of –0.09 to 0.31).
The frequencies of the CRI of cranium and sacrum differ significantly (MORAN AND GIBBONS, 183-190).

SOMMERFELD ET AL. (2004) investigate the intra- und interexaminer reliability of the palpation of the primary respiratory rhythm (PRM) which is simultaneously assessed by two osteopaths on 49 asymptomatic test persons.
The breathing frequencies of the examiners and of the test persons are registered as well as the onset of the phases of flexion and extension sensed by the osteopaths.

The PRM frequency, mean duration of the flexion phase and mean ratio of flexion to extension phase are evaluated by means of a four factor ANOVA variance analysis, the reliability of the different examiners is described by 95% confidence intervals and possible interactions between the palpation outcomes and the breathing frequency of test persons and examiners is assessed by analysis of covariance.
SOMMERFELD concludes that there are dependencies of the sensed PRM of the breathing frequency of the examiners, but not with the test persons, as well as a poor reliability of the palpations (SOMMERFELD ET AL., 4-10).

The results of these studies can be summarised in one sentence:
The more strictly the performance of a test is defined and the less ambitious the fundamental question is, the higher its reliability will be.

Also the following results have to be considered for similar investigations:
The examiner's experience is considered to have an influence on the reliability tests in some but not all studies.
Reliability of palpatory results is highly depending on body regions. Statements about one range of themes can not be generalised.
An environment supporting concentrated work, can influence the results in a positive way.

In my opinion, it has to be considered, that in many cases therapy is successful even in spite of different results of osteopathic diagnosis, indicating that different therapy concepts may lead to success.

3. Methodology

3.1. Basicals of the Investigation

3.1.1. Procedure of the Investigation

I proposed that 20 osteopaths should perform an abdominal global listening test (cf chapter 2.2.3) on 15 test persons. In spite of a long-term and multi-stage information process the proposed number of osteopaths could not be attained.

In a first stage, approximately two months in advance of the study, the readiness for taking part was ascertained in telephone and personal talks with Viennese osteopaths. In a second stage a written information about the date and the planned procedure of the tests was sent to those 31, who had given an affirmative answer.

Two weeks in advance of the fixed date these osteopaths were contacted again on phone, to confirm their coming.

Since the number of negative replies was unexpected high, it was attempted to motivate other osteopaths from Vienna and the surrounding districts for a participation in the study. In the end, after more negative replies only 14 osteopaths took part.

An unchanged or at least similar condition of the test persons for all therapists is of high importance for the study. Therefore, I refrained from a split of the investigation.

The tests took place on 2006-04-25.

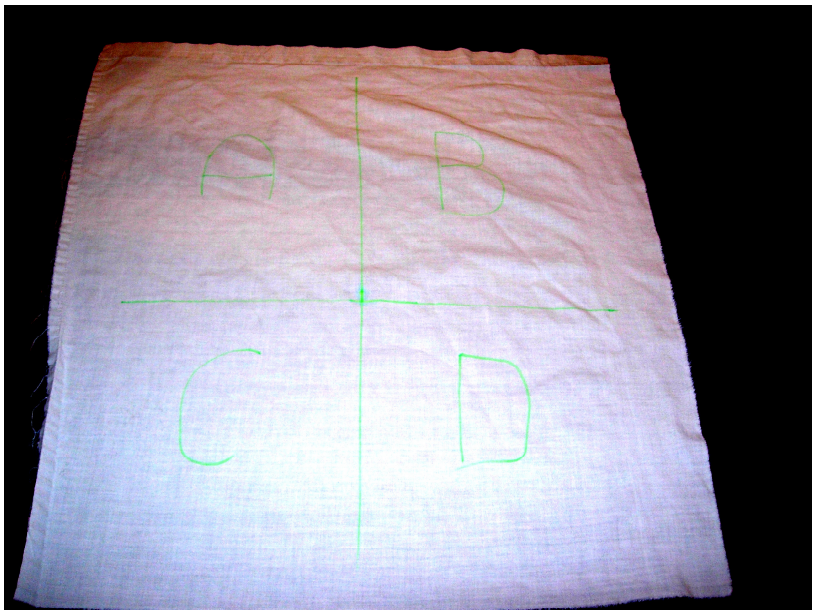
First of all, the osteopaths were lead through the therapy rooms and informed about the planned procedure. Additionally, the test persons got an introduction into the procedure.

Subsequently, data about the osteopaths' experience and favourite therapy methods were collected and identification numbers were assigned to them. The therapists queued in the order fixed in advance and entered the therapy rooms one after another.

The test person was already lying in supine position on the therapy bed, with his belly covered by a cloth which was marked with four quadrants (letters A - D, cf. Ill. 1), when the therapist arrived.

The cloth had been arranged by the conductor of the investigation at the beginning of each run.

The therapist did neither know the test persons nor their health problems, and additionally he could not notice if there were obvious problems, since the body was covered and the test person did not move. By that, a prepossession of the testing osteopaths should be excluded.



III.1: Cloth with quadrants.

The cloth was arranged in this way that the crossing of the quadrants laid over the navel with the horizontal line parallel to the spina eliaca ant. sup. (anatomical asymmetries in this connection were not considered).

In order to prevent muscle tensioning, eventually brought about by the length of the procedure, which might influence the results, the test person was asked to raise his pelvis before each test (cf. III. 2). Additionally, by this exercise it was expected to resolve possibly occurring therapy effects.

III.2: Resolving of muscle tensions.

Subsequently, the therapist laid his hand frontal medial onto the crossing point of the quadrants and tested for five seconds (cf. Ill. 3).
The time was controlled by independent persons, who also noted on a data sheet to which quadrant the therapist's hand was lead (data sheet cf. annex 2). Additional to the quadrants "A"-"D" for the horizontal movements, "O_T", characterizing a vertical movement only and "O", standing for no movement were the possible answers. This procedure was repeated by the other osteopaths.

Ill.3: Abdominal local listening.

3.1.2. The Test Persons

The 15 test persons were recruited from my friends and patients.
Exclusion criteria were large operation scars in abdomen and thorax and overweight. Exceptions were small scars after appendectomy or arthroscopical operations. Though, during the acquisition of the test persons it was taken care that these did not outnumber the others.
The line of overweight was drawn by the estimation:

Body weight – body height+100 >10

The 15 test persons, among them seven women and eight men, were between 20 and 45 years old. Four of them had had an appendectomy, two a navel piercing and one an arthroscopical operation.

The test persons were informed about the course of the investigations and it was stressed that they would not be treated during the tests. They took part on their free will and did not get any financial compensation for their participation.

3.1.3. The Osteopaths

It was planned that each osteopath taking part in this examination should have a similar practical experience, finishing the osteopathic education in the Vienna International School for Osteopathy (WSO) at the same year.
Due to the low number of osteopaths willing to take part in this study, this proposal could not be fulfilled. Therefore, the experience of the osteopaths was evaluated by the question whether abdominal local listening was used “routinely”, “often”, “rather seldom” or “rarely”. Since all osteopaths declared to use this test routinely, only the final year of the osteopathic education was relevant for the classification.
Additionally, the preferential body structure the osteopaths work on most often during therapy (visceral, cranio-sacral, structural or fascial) was collected. These data should help to assess influences of the experience in a certain field on the results. Also in this case, *all osteopaths* declared to have no preferences.

The therapists entered the therapy room one by one and performed the test on the test person. Their results were noted by an external person and kept hidden to other osteopaths. Therefore, no influence by other therapists could happen. This procedure was also followed in rooms where two test persons were laying.

The order of the osteopaths was fixed in a randomised way in advance of the study (cf. chapter 3.1.4.).

3.1.4. Objectifying of the Procedure

Preparation of the Osteopaths

In order to eliminate influences of accustoming difficulties of the osteopaths, the data diagnosed on the first two test persons were excluded from the evaluation, comparatively.
The first three osteopaths were asked to repeat their test on each subject after the local listening of the other therapists in order to be able to recognize influences of nervous tensions of the examiners. By the comparison of these results possibly a therapy effect can be evaluated or the intraexaminer reliability can be calculated,

respectively. A high agreement points toward a high intraexaminer reliability, whereas in the case of low agreements a treatment effect must be considered and researched in detail.

Preparation of the Test persons

It was considered to start the investigation with three non-osteopaths in order to familiarize the test persons to the situation. Their unawareness of this test should ensure that treatment effects which might influence the results were avoided. This topic could not be realized due to the lack of volunteers, but by the elimination of the data sets of the first three osteopaths for the evaluation, as described above, an appropriate phase for the reduction of nervous tensions should be guaranteed.

Randomisation of the Test Sequence

According to the literature, manual diagnosis techniques do not only act unidirectional. This means that information is not only transferred from the patient to the therapist, but also a therapy effect can be observed after a diagnostic touch. This effect is said to be differently high depending on the ability of the diagnostician to avoid it. Besides, in the literature (RUSSEL, 1983 and GOLDSTEIN, 1978) this therapy effect is only used for explanations of poor agreements but not yet proven. By a variation of the test sequence personal and temporal effects might be distinguishable, supposing there is a systematic in the test results.

Since five therapy beds are used at a time, due to practical reasons the sequence is only changed after five test persons resulting in three different sequences.

The test sequence was determined by means of random numbers:

Two random numbers between 0 and 20 are calculated.

One is related to the first position in the succession; equal to the identification number (ID) of the first osteopath to perform the test, the other is used as increment. The sum of the first position and the increment results in the identification number of the second osteopath. In the case that the ID would be >20, 20 is subtracted.

The sequences were inspected and those with duplicate values were excluded.

An example for the calculation can be seen in table 1: The first random number rn1 is 20 (first osteopath, position 1), the second random number rn2 (increment) is 11. For the determination of the second osteopath, his ID is calculated by the addition of 11 and 20. Since the result 31>20, 20 is subtracted. Therefore, O11 is the next

osteopath. The ID of the third osteopath is calculated by 11+11. Since also 22>20, 20 is subtracted again and O2 is the third to perform the test (followed by O13, O4,…)

Position		Example	ID
1	ID1 = rn1* (1<rn1<=20)	20*	O20
2	ID2 = rn1* + rn2** (1<rn2<=20)	20+11**=31 31>20=> 31-20= 11	O11
3	ID3= ID2 + rn2**	11+11**= 22 22>20=> 22-20= 2	O2

Table 1: Calculation of the test sequence (rn… random number).

The succession of the osteopaths was determined in advance of the tests on the basis of 20 osteopaths. Since only 14 osteopaths participated in the test the numbers which were not used were stricken out and the succession was maintained under omission of the missing osteopaths. Thus and due to the randomisation, two osteopaths tested ten test persons twice and five only once, another five osteopaths five test persons twice and ten once. Seven therapists performed only one test on all 15 test persons. In total, 255 tests were performed.

3.1.5. The Data Sheet

In the data sheet (cf. Annex 2) the identification numbers of the test person and of the osteopaths as well as the quadrants resulting from the test were collected in the order of their appearance.

Additionally, scars on the thorax and abdomen (after appendectomy and arthroscopy operations) as well as body piercings and their positions were registered, proceeding on the assumption that scars which are not cleared might influence the local listening. Also here it has to be remarked that large operation scars are an exclusion criterion.

3.2. Evaluation of the Results

The data gained during the investigation were collected in a data bank (MS Access 97) and Cohen's Kappa (κ) was calculated for the agreements of all osteopaths one with another and *all quadrants together*. The fundamental equation for the calculation was taken from SACKET ET AL., 1991. The computer-aided calculations were directly performed in the data bank by means of macros and queries.

For the explanation of the procedure, in table 2 a schematic 5x5 contingency table of the test results is shown. The numbers of agreements (O_{ii}) between the two examiners in the diagonal are marked dark blue, the number of differing results light blue. The orange cells represent calculated values, where C_i are the column sums and R_i the row sums.

		Examiner 1					
Results		A	B	C	D	0	Row totals (R)
Examiner 2	A	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	R_1
	B	O_{21}	O_{22}	O_{23}	O_{24}	O_{25}	R_2
	C	O_{31}	O_{32}	O_{33}	O_{34}	O_{35}	R_3
	D	O_{41}	O_{42}	O_{43}	O_{44}	O_{45}	R_4
	0	O_{51}	O_{52}	O_{53}	O_{54}	O_{55}	R_5
Column totals (C):		C_1	C_2	C_3	C_4	C_5	N

Table 2: 5x5 Contingency table

The number of agreements *expected on chance* is calculated for each cell by the formula:

$$E_{ii} = \frac{R_i \cdot C_i}{n}$$

and thus the probability of an agreement on chance p_E can be expressed by:

$$p_E = \frac{\left(\sum_{i=1}^{l} E_{ii}\right)}{n}$$

The probability of the *actual agreement* p_0 is calculated by

$$p_0 = \frac{\left(\sum_{i=1}^{l} O_{ii}\right)}{n}$$

from the diagonal cells.

Kappa (κ) usually is expressed in the following standardized way:

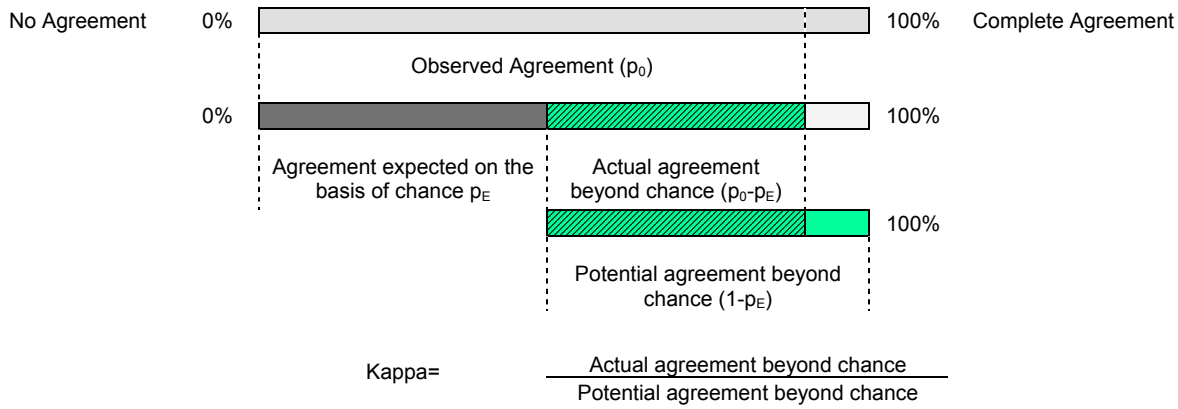
$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

The results of κ were interpreted with the degrees of agreement commonly used in the later literature (LANDIS AND KOCH, 1977, 159-174):

κ < 0.20	poor
0.20 < κ < 0.40	fair
0.40 < κ < 0.60	moderate
0.60 < κ < 0.80	substantial
0.80 < K < 1.00	almost perfect

According to FJELLNER ET AL., 1999, 511-516 κ-indices higher than at least 0.4 are considered as indicator for an acceptable interobserver-reliability.

In WOODWORD, 1999 a clear idea of the kappa-index, as well as its context with the agreements is given (cf. Ill. 4):



Ill.4: Explanation of the meaning of the κ-index.

In order to show possible dependences of single positions (quadrants) the results of two therapists each were calculated and classified *for each single quadrant* according the following 2x2 contingency table ((ROSNER, 1986) and (FEINSTEIN, s.a.)),additionally (cf. Table 3).

Examiner 2 (Q_{ii})		Examiner 1 (Q_i) P	Examiner 1 (Q_i) A
	P	Q Number of agreements on presence	R Number of disagreements
	A	S Number of disagreements	T Number of agreements on absence

Table 3.: 2x2 Contingency table, Q_i =A, B, C, D, 0, …

In this simplified case the formulas for the calculation of kappa are:

$$N = Q + R + S + T$$

$$P_0 = \frac{Q + T}{N}$$

$$P_e = \left[\frac{Q + R}{N}\right]\left[\frac{Q + S}{N}\right] + \left[\frac{R + T}{N}\right]\left[\frac{S + T}{N}\right]$$

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

Subsequent to these fundamental calculations the κ-indices for therapists with similar experience classified in two groups were calculated.

The final year of the osteopathic education was used for the classification, since no more accurate distinction by the question about the frequency of use of local listening could be gained. The osteopaths were sorted by their final year at the WSO. The median demarked the line between osteopaths with more and less experience. The group with osteopaths finishing their studies between 1998 and 2001 was classified as “higher experienced”, the group with those finishing between 2002 and 2005 as “less experienced”.

In order to be able to recognize therapy effects, in a further step κ was calculated for *the test positions* independently from the individual osteopaths.

Additionally, the estimated 95% - confidence intervals for κ were calculated in all cases mentioned above.

$\kappa \pm 1.96 \cdot \hat{s}e(\kappa)$ with

$$\hat{s}e(\kappa) = \sqrt{\frac{1}{(1-p_e)^2 \cdot n} \cdot \left\{ p_E^2 + p_E - \sum_{i=1}^{l} \frac{R_i \cdot C_i}{n^3} \cdot (R_i + C_i) \right\}}$$

J. L. FLEISS suggests an additional statistical valuation test, in which the null hypothesis H_0: κ=0 (ie. result on the basis of chance) is compared with the alternative hypothesis H_1: κ ≠ 0 (result not on the basis of chance) (FLEISS, 1981, 321).

For that purpose, the test statistic $Z = \left(\frac{\kappa}{\hat{s}e(\kappa)} \right)^2$ with the critical value $\chi^2_{(1-\alpha,1)}$ from the critical values of the χ^2- distribution is compared for one degree of freedom.
(Tables: FISCHER AND YATES, 1982, 146).

4. Results

4.1. Basicals of the Evaluation

4.1.1. The Osteopaths

In table 4 the final years of the osteopathic education at the WSO and the affiliation to the two groups (“more experienced” and “less experienced”) of the osteopaths are collected. The question about the frequency of the usage of the abdominal local listening was answered with “regularly” by all osteopaths, and also the question about the favourite treatment method was not answered in a categorized way as intended. The therapists use a mixture according to the necessity.

Osteopath	Final year	Group
O3	1998	
O8	1998	
O9	1999	
O4	2000	More experienced osteopaths
O13	2000	
O5	2001	
O14	2001	
O12	2002	
O1	2004	
O2	2005	
O6	2005	Less experienced osteopaths
O7	2005	
O10	2005	
O11	2005	
Median	2001.5	

Table 4: Experience of the osteopaths (affiliation by the final year of education at the WSO).

4.1.2. Data of the Test persons and Raw Data

The raw data can be seen in table 5. Additionally, possible disturbances (little scars) and their reasons are described. It can be seen from these data, that the dispersion of the results is high. Except in the case of test person P14, the little scars can not be brought in connection with the results of the tests. In the last two rows of the table “A*“ demarks the number of agreements of the scar positions and the actual tested position (exact quadrants) and A** the number of agreements of scar position and tested positions after the splitting of double quadrants. That means, if double quadrants were postulated, as for example AB, in the latter case A as well as B was considered.

Test-position	Test person														
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
1	O	C	A	C	AB	CD	O	AB	O	O	A	O	AC	C	AC
2	B	B	C	C	CD	C	B	D	O	OT	C	A	C	C	A
3	CD	O	OT	OT	CD	D	C	D	CD	AB	OT	C	C	C	AC
4	B	C	B	B	O	A	O	C	D	BD	B	AB	A	CD	A
5	D	D	O	D	AC	B	C	A	B	O	B	B	A	AB	A
6	B	D	A	AB	A	C	B	C	OT	A	D	B	A	D	C
7	B	B	O	OT	A	C	C	O	C	A	D	D	B	D	B
8	D	B	A	B	D	A	AC	O	O	O	B	B	O	A	CD
9	C	C	C	C	A	BD	D	A	AC	C	C	C	O	D	B
10	OT	C	O	OT	A	C	C	D	D	C	B	C	A	C	CD
11	C	O	C	C	O	C	B	A	D	B	C	B	OT	B	B
12	C	C	B	C	A	A	AB	C	A	AC	D	B	OT	B	D
13	B	D	C	CD	C	B	A	O	B	D	D	B	CD	CD	B
14	O	A	D	D	C	D	B	AB	C	A	B	A	D	O	B
15	O	O	A	C	A	O	O	AB	C	O	C	C	O	O	AC
16	B	D	B	D	B	C	B	CD	B	CD	A	C	A	C	A
17	D	B	CD	O	CD	B	O	AB	CD	B	A	O	B	C	C
Scar Reason	C Ap		O NP		C Ap	C Ap							O NP	C+D A	C Ap
A*	3		4		2	6							5	11	2
A**	4		4		6	7							5	11	7

Table 5: Raw data of the tests and comparison with the quadrants where little scars are situated. C+D...scars in quadrant C as well as D, Ap...Appendectomy, NP...Navel piercing, A...Arthroscopy.

Noticeable are also repeatedly occurring sequences of quadrants tested in succession. For example, three times quadrant “D” was tested on test person P1 after the result “’B”, if “CD” is interpreted as “D” even four times.

In the following table 6 sequences of two quadrants, which occur more often than twice and their relative frequency are described.

Test person	Sequence	Number	Rel. Frequency %
1	BD	4	25.0
4	CC	3	18.8
4	CD	3	18.8
6	AB	3	18.8
6	CC	3	18.8
8	AD	3	18.8
9	OC	3	18.8
11	BC	3	18.8
12	BB	3	18.8
14	CC	4	25.0
15	AA	5	31.3
15	BC	3	18.8
15	BD	3	18.8
15	CA	3	18.8
15	CB	3	18.8
15	DB	3	18.8

Table 6: Sequences of two quadrants occurring more often than twice in the tests at the same test person.

As can be read from table 6, also sequences with different quadrants can be observed. There is no point in a detailed inspection of the contingency due to the small sample size. Since there are maximum 15 equal quadrants (“CC”) in the whole dataset (i.e. 240 possible sequences of two), it can be estimated, that the maximum relative frequency of equal sequences is 6.25%. The relative frequencies in table 6 are approximately three times higher, indicating that the result might occur not only by chance.
Thus, the repeated occurrence of these sequences of two different quadrants might be interpreted as a consequence of a body rhythm (analogous inspir or exspir).

In table 7 the raw data are summarized. Again, the quadrants tested most frequently are marked with blue. The positions of the scars, as well as the frequencies how often the quadrants where they are located were dated can be compared, too.

Quadrant	Number of the single traced quadrants														
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
A	0	1	4	0	6	3	1	3	1	3	3	2	5	1	4
B	6	4	3	2	1	3	5	0	3	2	5	6	2	2	5
C	3	5	4	6	2	6	4	3	3	2	4	5	2	6	2
D	3	4	1	3	1	2	1	3	3	1	4	1	1	3	1
O	3	3	3	1	2	1	4	3	3	4	0	2	3	2	0
OT	1	0	1	3	0	0	0	0	1	1	1	0	2	0	0
AB	0	0	0	1	1	0	1	4	0	1	0	1	0	1	0
AC	0	0	0	0	1	0	1	0	1	1	0	0	1	0	3
CD	1	0	1	1	3	1	0	1	2	1	0	0	1	2	2
BD	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Scar	C		O		C	C							O	C+D	C
A*	3		4		2	6							5	11	2

Table 7: Number of the single traced quadrants.

Since it is unlikely that the movement is running along the dividing line between two quadrants (e.g. AB) in so many cases and it is more likely that the examiner can not trace little deviations from it, the combined quadrants were split in their single components (A and B) and the frequencies were counted again. These data can be seen in table 8. In the case of combined quadrants both quadrants were taken into consideration, resulting in a higher total number. Thus, also the relative frequencies are presented.

Quadrant	Number of the single quadrants (under consideration of movements along dividing lines)														
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
A	0	1	4	1	8	3	3	7	2	5	3	3	6	2	7
B	6	4	3	3	2	4	6	4	3	4	5	7	2	3	5
C	4	5	5	7	6	7	5	4	6	4	4	5	4	8	7
D	4	4	2	4	4	4	1	4	5	3	4	1	2	5	3
O	4	3	4	4	2	1	4	3	4	5	1	2	5	2	0
Sum	18	17	18	19	22	19	19	22	20	21	17	18	19	20	22
Scar	C		O		C	C							O	C+D	C
Reason	Ap		NP		Ap	Ap							NP	A	Ap
A** (abs)	4		4		6	7							5	13	7
A** (rel)	0.22		0.22		0.27	0.37							0.26	0.65	0.32

Table 8: Frequency of the different results after the dissolution of combined quadrants demarking movements along a dividing line between two quadrants.

Also from these data and the comparison of the position of scars with the results an influence of scars on the result can only be observed on the test persons P6 and P14. In A** (rel) the relative frequency and in A** (abs) the absolute number of agreements between the scar position and the test results are presented.

4.2. Interexaminer Reliability

The one-by-one agreements of the osteopaths in diagnosis with the abdominal local listening tests will be characterised by means of κ-indices in this chapter.
By exclusion of different data sets various influences on the results will be precluded. These results of the different approaches will be presented in the following chapters (chapters 4.2.1.1. and 4.2.1.2.) and finally be compared in an own subchapter at the end (chapter 4.2.1.3.).

4.2.1. Effects of Accustoming and Tensing of the Test Persons

In the first subchapter (4.2.1.1.), the tests of the first three osteopaths will not be included in the calculations, as initially intended. In the second subchapter (4.2.1.2.), for additional information, the tests of the osteopaths, who tested a second time will be removed from the dataset before the calculations. The differences between the results shall be used to work out influences due to accustoming and tensing effects on the test persons.
In addition, each of these calculations was performed with and without the results gained from the tests on the first two test persons. These data are used to find out, if an accustoming phase for the osteopaths is necessary.

Tabular overviews of the one-by-one agreements between all osteopaths will be given by means of the κ-indices. Cells above the diagonal which are not marked with colours contain values, which come about by chance, light yellow cells contain values indicating poor reliability, and light green ones a fair reliability. Values below the diagonal would be redundant and thus were omitted for a clearer arrangement.

4.2.1.1 Results without the Data of the First Three Tests

Only the results between the 4th and 17th test position *obtained from all test persons* were used for the evaluation of the data in this case. That means that only the second tests of the three osteopaths performing the first three and last three tests were taken into account. Thus, the osteopaths marked with an asterisk in table 10 delivered data on a position between 4 and 17, the others between 4 and 14.

The reason for the elimination of the first three results on each test person was to preclude influences of an initial tensing due to accustoming problems.

Mean values, the 95%- confidence intervals (95% CI), standard deviations and medians of the κ-indices of all comparisons, as well as the same data only for positive κ-values are given in table 9, the single data in table 10.

κ (4-17)	Mean	95% CI	SD	Median
all κ	0.0048	±0.019	0.0952	0.0000
positive κ	0.081	±0.019	0.062	0.063

Table 9: Descriptive data of κ for the osteopaths calculated from the results gained at the positions 4-17.

In total, these results of table 9 show an only slight better agreement of the osteopaths than the expected agreement on basis of chance. If only the results beyond the agreement on chance are considered, the mean reliability is only poor. Nevertheless, in table 10 can be seen that for example the pairs of osteopaths O6 and O9, O7 and O12, as well as O7 and O14 did sense the same at the same test person more often, reaching fair agreements.

	O	Examiner 2													
		1*	2	3*	4*	5*	6	7	8	9	10	11*	12	13*	14*
Examiner 1	1*		-0.12	0.03	0.03	-0.09	0.06	-0.13	-0.02	0.02	0.03	-0.16	-0.13	-0.04	-0.05
	2			-0.23	0.04	-0.10	-0.07	0.08	-0.05	-0.05	-0.07	0.06	0.10	0.17	-0.03
	3*				-0.07	0.18	0.06	-0.02	0.14	0.05	-0.05	-0.06	0.03	-0.04	-0.03
	4*					-0.06	-0.03	-0.08	-0.02	-0.17	0.15	-0.09	0.09	-0.09	-0.21
	5*						0.11	0.12	-0.03	0.06	0.07	-0.04	0.03	-0.15	-0.01
	6							0.03	-0.03	0.31	0.10	-0.08	-0.09	0.03	-0.08
	7								0.12	-0.10	-0.05	0.04	0.21	0.07	0.20
	8									0.05	-0.01	0.07	0.11	0.02	0.02
	9										0.11	0.00	0.08	0.04	-0.04
	10											0.15	-0.03	-0.08	0.01
	11*												0.06	-0.03	0.10
	12													0.02	0.02
	13*														-0.01
	14*														

Table 10: κ-indices for all pairs of osteopaths calculated from all results of the positions 4 - 17.

By this elimination of the first tests on each test person only the accustoming of those osteopaths performing these three tests is considered. Thus all results of the tests gained from the first two test persons were eliminated from the raw data set, too. The results after the calculations of the κ-indices with the remaining data are presented in table 11.

		Examiner 2													
	O	1*	2	3*	4*	5*	6	7	8	9	10	11*	12	13*	14*
Examiner 1	1*		-0.14	0.05	0.06	-0.05	0.10	-0.11	0.01	0.04	0.06	-0.25	-0.13	-0.04	0.00
	2			-0.15	0.05	-0.07	-0.10	0.10	0.02	0.01	-0.10	0.06	0.13	0.08	-0.04
	3*				-0.08	0.15	0.03	-0.03	0.04	-0.05	-0.06	-0.05	0.05	0.02	-0.04
	4*					-0.13	-0.04	-0.13	-0.03	-0.17	0.18	-0.10	0.07	-0.09	-0.21
	5*						0.13	0.09	-0.10	0.03	0.05	0.01	0.01	-0.13	0.03
	6							0.04	-0.09	0.29	0.12	-0.08	-0.06	-0.03	-0.10
	7								0.14	-0.08	-0.03	0.06	0.15	0.10	0.21
	8									-0.08	-0.05	0.09	0.14	0.07	0.03
	9										0.06	0.00	0.14	0.08	-0.04
	10											0.18	-0.01	-0.11	0.01
	11*												0.06	-0.02	0.10
	12													0.06	0.00
	13*														-0.01
	14*														

Table11: κ-indices for all pairs of osteopaths after exclusion of the test results on the first two test persons from the results of the positions 4-17.

By the elimination of the results on the Test persons P1 and P2 the number of results characterising agreements on chance ($\kappa \leq 0$) decreases, indicating an accustoming effect of the therapists. Nevertheless, the number of κ-values showing fair reliability is decreased due to agreements also in the tests on the first two subjects, too. In table 12 mean values, medians, standard deviations and the 95%-confidence intervals (95% CI) for all κ and only for positive κ are listed.

κ (4-17) without P1 and P2	Mean	95% CI	SD	Median
all κ	0.0047	±0.020	0.0980	0.0069
positive κ	0.081	±0.018	0.0597	0.065

Table 12: Descriptive data of κ for all osteopaths after exclusion of the datasets of the first two test persons from the results of the positions 4-17.

As can be seen from the tables 11 and 12 no significant change of the results can be deducted.

4.2.1.2 Results without the Last Three Tests (Repeated Tests)

Since tensioning might also arise from the long lying of the test persons during the test procedure (also in spite of the lifting of the pelvis), this time the calculation of the κ-indices was performed without the results of the last three measurements. Thus, all data from the positions 1-14 are taken into account. Mean values, medians, standard deviations and 95%-confidence intervals (95% CI) for all and only positive κ, respectively, are shown in table 13. The κ-indices for all pairs of osteopaths are presented in table 14.

κ (1-14)	Mean	95% CI	SD	Median
all κ	0.0048	±0.019	0.0952	0.0052
positive κ	0.078	±0.021	0.0697	0.067

Table 13: Descriptive data of κ for all osteopaths from the results of the positions 1-14.

These results are similar to those presented in the chapter above. The data in table 14 show that more better agreements arise, if the results of the tests repeated by the first three osteopaths are not considered in the evaluation compared to the calculation without the results of the first three tests.

		Examiner 2													
	O	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Examiner 1	1		-0.12	-0.08	0.04	-0.07	-0.01	-0.17	0.00	-0.07	-0.13	-0.06	-0.26	-0.04	-0.05
	2			-0.12	0.00	-0.09	-0.07	0.08	-0.05	-0.05	-0.07	0.06	0.10	0.07	-0.04
	3				-0.11	0.01	-0.08	-0.03	0.04	0.05	0.03	-0.06	0.25	0.21	0.08
	4					0.01	-0.01	-0.01	0.02	-0.15	0.23	-0.06	-0.02	-0.05	-0.03
	5						0.03	0.08	-0.04	0.08	0.08	0.07	0.07	-0.03	0.05
	6							0.03	-0.03	0.31	0.10	-0.07	-0.09	-0.01	-0.09
	7								0.12	-0.10	-0.05	0.02	0.21	0.08	-0.02
	8									0.05	-0.01	0.01	0.11	0.01	0.01
	9										0.11	-0.01	0.08	0.07	-0.03
	10											0.07	-0.03	-0.06	0.01
	11												0.03	-0.03	-0.03
	12													0.10	0.05
	13														0.05
	14														

Table 14: κ-indices for all pairs of osteopaths calculated from the results of the positions 1-14.

Without the results of the first two test persons (P1, P2), which were excluded from the dataset in order to evaluate possible accustoming effects for the therapists, the κ-indices presented in table 15 were calculated.

		Examiner 2													
	O	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Examiner 1	1		-0.13	-0.10	0.07	-0.02	0.02	-0.14	0.03	-0.07	-0.12	-0.11	-0.27	-0.05	0.00
	2			-0.15	0.00	-0.06	-0.10	0.10	0.02	0.01	-0.10	0.08	0.13	0.08	-0.04
	3				-0.12	0.00	-0.04	0.00	0.03	0.03	-0.10	-0.04	0.32	0.28	0.11
	4					-0.04	-0.01	-0.04	0.03	-0.17	0.28	-0.13	-0.04	-0.03	-0.01
	5						0.04	0.04	-0.11	0.05	0.07	0.03	0.04	0.02	0.08
	6							0.04	-0.09	0.29	0.12	-0.07	-0.06	-0.01	-0.11
	7								0.14	-0.08	-0.03	-0.01	0.15	0.11	-0.04
	8									-0.08	-0.05	0.01	0.14	0.03	0.01
	9										0.06	0.01	0.14	0.10	-0.03
	10											0.11	-0.01	-0.02	0.01
	11												-0.01	-0.02	0.03
	12													0.13	0.03
	13														0.01
	14														

Table 15: κ-indices for all pairs of osteopaths after exclusion of the data gained on the first two test persons (positions 1-14).

Also in this case a flattening of the indices occurs. This means that the κ-indices indicating a poor reliability come out more often, whereas worse and better results are more seldom. Equal to the outcomes of the last chapter, there is evidence that some osteopaths need an accustoming phase and some other do not.

Mean values, medians, standard deviations and 95%-confidence intervals (95% CI) for all and only positive κ, respectively, are shown in table 16.

κ (1-14) without P1 and P2	Mean	95% CI	SD	Median
all κ	0.0047	±0.020	0.0980	0.000
positive κ	0.083	±0.024	0.080	0.059

Table 16: Descriptive data of κ for all pairs of osteopaths (positions 1 –14) after exclusion of the first two test persons.

Generally seen, also this overview of the results indicates that the agreement between all osteopaths is only slightly better than expected on the basis of chance. The slightly higher mean value of the positive κ-indices might arise from the fact, that more therapists take profit from the accustoming phase simulated by the exclusion of the data of the two first test persons. By this, the results of all therapists are treated same and the effect is more distinct.

The extent of this effect is extremely small compared to the other possible influences.

4.2.1.3 Summary

If the descriptive data of the four tables 9, 12, 13 and 16 are compared (cf. table 17), only slight differences of the four cases presented above can be observed.
Further comparisons by test statistics are not necessary in this case.

	Positive and negative κ				Positive κ			
Case	Mean	95% CI	SD	Median	Mean	95% CI	SD	Median
4-17	0.0048	±0.019	0.0952	0	0.081	0.019	0.062	0.063
4-17*	0.0047	±0.020	0.0980	0.0069	0.081	±0.018	0.0597	0.065
1-14	0.0050	±0.019	0.0916	-0.0052	0.078	±0.021	0.0697	0.067
1-14*	0.0064	±0.020	0.1008	0	0.083	±0.024	0.080	0.059

Table 17: Descriptive data of κ for the four evaluations basing on different data sets.

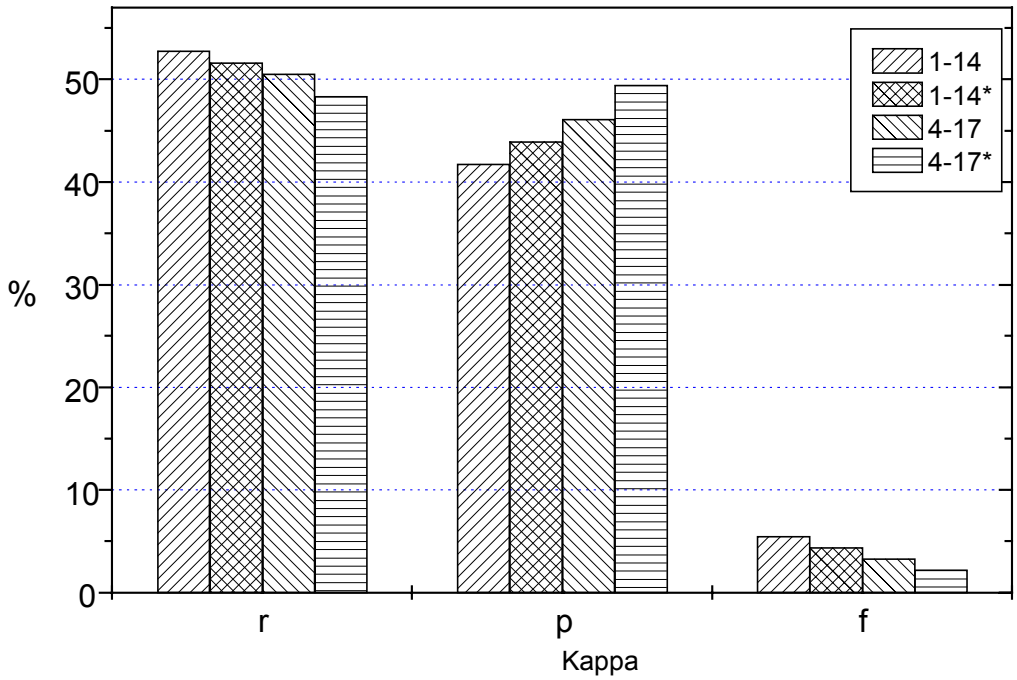
The only possible interpretation of these results is, that this test is not universally valid in the way as it was performed.

Nevertheless, agreements between pairs of therapists beyond those expected on chance, occurring in approximately 50% of all cases, indicate that these osteopaths did measure the same at the same test persons more often. The best agreements have to be classified as fair agreements.

The numbers of the agreements classified with the degrees of agreement commonly used in the literature can be read from illustration 5 and table 18.
κ-indices with $\kappa \leq 0$ are marked with an r (random), κ-indices with $0<\kappa<0.2$ with a p (poor) and κ with $0.2<\kappa<0.4$ with an f (fair).

	1-14	1-14*	4-17	4-17*
Random (r)	48	47	46	44
Poor (p)	38	40	42	45
Fair (f)	5	4	3	2

Table 18: Absolute numbers of the degrees of agreement for the four differently evaluated data basis. *...without test person P1 and P2.



Ill. 5: Comparison of the relative frequencies of the degrees of agreements for the four differently evaluated data basis (*...without test person P1 and P2).

As can be read from table 17 and illustration 5, more differing measurements happen in the first three tests (position 1-14) compared to the last three tests (position 4-17). Consequently, in the case 4-17 a higher number of agreements beyond the expected agreements on the basis of chance ($\kappa>0$) can be observed.

Without the consideration of the results of the first two test persons (case 4-17*), the number of agreements is highest. Since the degree of agreement is also reduced in the cases of a positive value of κ (which can be observed at the lower number of cases with fair reliability), no effect on the mean values can be measured.

Thus, there is evidence that accustoming effects of therapists and test persons influence the result. Compared to the low agreements these factors are not of high importance.

4.2.2. Results Evaluated Considering the Order of Tests.

In order to find out possible influences of therapy effects, κ-indices were calculated for all pairs of test positions *(independently of the testing osteopaths!)*.

For example, the 15 results gained from the tests on the first test positions on the 15 test persons were compared with all results of tests on the second (3rd, 4th, …, 17th) test position by the calculation of κ-indices. The individual results are summarised in table 19.

In the case of therapy effects, more remote positions should result in lower κ-indices (or decrease from left to the right), because the appearing of a new pattern should result in fewer agreements. Though, this effect can only be observed under the prerequisite of a better agreement between the therapists and the inset of the effect after a similar number of tests.

Position	Po1	Po2	Po3	Po4	Po5	Po6	Po7	Po8	Po9	Po10	Po11	Po12	Po13	Po14	Po15	Po16	Po17
Po1		0.09	0.03	0.03	-0.04	-0.02	-0.11	0.08	-0.01	0.01	-0.06	0.04	-0.07	0.03	0.29	0.06	0.24
Po2			0.14	-0.02	-0.07	0.03	0.03	0.03	-0.04	-0.02	0.14	-0.15	-0.03	-0.02	-0.03	0.18	0.06
Po3				-0.08	0.00	-0.11	0.03	-0.06	-0.10	0.20	-0.06	-0.13	-0.10	-0.02	0.10	0.04	0.11
Po4					0.05	0.06	-0.11	0.05	-0.07	0.17	-0.04	0.15	-0.10	-0.10	-0.07	0.11	-0.10
Po5						0.04	-0.04	0.11	-0.06	0.18	-0.03	-0.09	0.13	-0.05	-0.06	0.20	-0.01
Po6							0.27	0.00	-0.04	0.04	0.04	0.10	0.14	-0.05	0.04	0.18	-0.06
Po7								-0.03	0.04	0.21	-0.08	-0.04	0.11	0.03	0.00	-0.06	-0.02
Po8									-0.02	0.05	-0.09	0.02	-0.03	-0.12	0.05	-0.18	-0.03
Po9										0.05	0.25	0.10	0.01	-0.09	0.19	-0.11	-0.13
Po10											-0.07	-0.05	-0.14	-0.08	-0.03	0.13	-0.06
Po11												0.18	0.02	-0.03	0.05	-0.08	-0.12
Po12													-0.01	-0.18	0.03	-0.14	-0.13
Po13														-0.03	-0.08	0.01	-0.12
Po14															0.16	-0.04	-0.08
Po15																-0.01	-0.01
Po16																	-0.04
Po17																	

Table 19: κ-indices of all pairs of positions.

54 values (59.3%) of the 91 values in the big black box in table 19, pointed out an agreement only on the basis of chance, 33 (36.3%) showed a poor agreement and four (4.4%) a fair one. The positions Po4-Po17 were used to guarantee a comparability with the data in the chapters above.

Also for the values in table 20, containing a descriptive summary of the results, only the last 14 positions (big black box in table 19) were considered.

κ (Po4-Po17)	Mean	95% CI	SD	Median
all κ	0.0012	±0.021	0.101	-0.030
positive κ	0.101	0.024	0.072	0.098

Table 20: Descriptive data for κ for the results of the positions Po4-Po17.

The lower mean and median values of all results, but higher values of the positive κ-indices compared to the ones presented in the chapter 4.2.1.3, give evidence that there exist influences of the order, which might base on equal results measured one after another.

These low values also indicate that the agreement between certain osteopaths has a higher influence than the order of tests.

The values from the diagonal in table 19, demarking subsequent positions, are extra specified in table 21 for a better overview.

Position	Kappa
Po1-Po2	0.09
Po2-Po3	0.14
Po3-Po4	-0.08
Po4-Po5	0.05
Po5-Po6	0.04
Po6-Po7	0.27
Po7-Po8	-0.03
Po8-Po9	-0.02
Po9-Po10	0.05
Po10-Po11	-0.07
Po11-Po12	0.18
Po12-Po13	-0.01
Po13-Po14	-0.03
Po14-Po15	0.16
Po15-Po16	-0.01
Po16-Po17	-0.04

Table 21: κ-indices for positions following the other.

The descriptive data summarising the data of table 21 are presented in table 22.

κ (Po4-Po17)	Mean	95% CI	SD	Median
all κ	0.043	0.054	0.101	0.016
positive κ	0.123	0.068	0.082	0.113

Table 22: Descriptive data of κ for the results in table 21.

Compared to the data in table 20 a mean agreement approximately 40 times higher can be observed (all κ), if only data of subsequent positions are considered. The agreement beyond chance (positive κ-indices) is higher than in the different comparisons of the individual osteopaths (cf. table 18), confirming that same quadrants were sensed more frequently in test positions following upon the last one than in more remote test positions.

In table 21 a maximum agreement between the 6th and 7th position can also be observed, which might indicate a later accustoming than expected.
Since, in this connection, there is no evidence of a trend in the following data (position 8-17) and since there are even more results demarking only a agreements on basis of chance after the 7th position, it can be concluded that other effects than accustoming have a higher influence on the result.
Considering the κ-indices of the six osteopaths who tested the test persons on the positions 6 and 7, it can be seen, that their total inter-individual agreement (agreement on all 15 test persons) is very low (cf. table 23) and therefore, not the reason for this result.

Test persons	Position 6	Position 7	κ (O_{pos6}-O_{pos7})
1-5	O8	O10	-0.01
6-10	O9	O7	-0.10
11-15	O12	O2	0.10

Table 23: Agreement of the results of the osteopaths working at the positions 6 and 7 (κ calculated from all 15 test results).

In table 21 also can be seen, that agreements beyond chance are gained especially in the first half of the test (position Po1- Po7), whereas after the 7th test position the agreements on the basis of chance predominate.
Even when a therapy effect, which should be noticeable in a chance of initial patterns to other patterns, can not be read from the data due to the high heterogeneity, a disturbing of the circumstances might be the reason for these observations.

Since there is only low probability that these effects occur exactly after the 7th test on each test person, separate evaluations of the positions 1-7 and 8-14 were performed and subsequently, the resulting κ-indices were compared as dependent variables by means of a t-test. The differences (cf. table 24) between the first and the second half of the investigation are significant on a level of significance of α=10, which was chosen due to the data characteristics (t=1.87; p=0.06).

	n	Mean	SD	Median
O1-14 (Po1-Po7)	56	+0.023	0.144	0.00
O1-14 (Po8-Po14)	56	-0.037	0.189	-0.25

Table 24: Descriptive data of κ for all osteopaths, grouped by the positions 1-7 and 8-14. No consideration of repeated measurements.

From these results - errors due to the heterogeneity of the results expected - it can be concluded, that the conditions did change during the investigation.

4.2.3. Results under Consideration of the Experience of the Therapists

The differences in the reliability between the results of individual therapists mentioned in the literature (e.g. MIOR ET AL., 1990 and BEAL AND PATRIQUIN, 1995) were also evaluated. The κ-indices of osteopaths grouped in therapists with higher experience and those with lower experience can be seen in table 25.

Less experienced osteopaths

		Examiner 2						
	O	3	4	5	8	9	13	14
Examiner 1	3		-0.07	0.18	0.14	0.05	-0.04	-0.03
	4			-0.06	-0.02	-0.17	-0.09	-0.21
	5				-0.03	0.06	-0.15	-0.01
	8					0.05	0.02	0.02
	9						0.04	-0.04
	13							-0.01
	14							

Osteopaths with more experience

		Examiner 2						
	O	1	2	6	7	10	11	12
Examiner 1	1		-0.12	0.06	-0.13	0.03	-0.16	-0.13
	2			-0.07	0.08	-0.07	0.06	0.10
	6				0.03	0.10	-0.08	-0.09
	7					-0.05	0.04	0.21
	10						0.15	-0.03
	11							0.06
	12							

Table 25: κ-indices for all pairs of osteopaths calculated from the results from all results of the positions 4-17, grouped by the experience of the osteopaths.

In table 26 mean values, medians, 95% - confidence intervals and standard deviation for the κ-indices are summarised.

	Mean	95% CI	SD	Median
Less experienced	-0.173	±0.0425	0.0934	-0.019
More experienced	-0.0016	±0.0458	0.1005	0.027

Table 26: Descriptive data for κ for all osteopaths (positions 4-17), grouped by the experience of the osteopaths.

From these data as well as from the numbers of agreements beyond the agreement expected on chance (cf. table 27) a possible influence of the experience can be deducted. Though, in a t-test no differences with statistical significance can be found (t=0.53, p=0.60).

	Less experienced		More experienced	
	n	%	n	%
Random	13	61.9	10	47.6
Poor	8	38.1	10	47.6
Fair	0	-	1	4.8

Table 27: Number of results grouped by the degrees of agreement
(Positions 4 –17, grouped by the experience of the therapists).

4.3. Intra-examiner Reliability

The data of the first three tests and those of the last three tests, which were performed by the same osteopaths, were used to calculate the κ-indices for the intra-examiner reliability (cf. table 28).

		2nd examination						
	O	1	3	4	5	11	13	14
1st examination	1	-0.15						
	3		0.04					
	4			-0.05				
	5				0.33			
	11					0.47		
	13						0.18	
	14							-0.05

Table 28: κ-indices of the intra-examiner reliability.

Also in this case, agreements which can not be interpreted as beyond chance (three of seven therapists, 42.9%) occur. Two agreements can be designated as poor (28.6%) and one as fair and moderate, respectively (14.3%).
Mean value, median, standard deviation and 95%-confidence interval (95% CI) for all κ-indices are shown in table 29.

	Mean	95% CI	SD	Median
κ_{Intra}	0.11	±0.209	0.226	0.04

Table 29: Descriptive data of κ for all osteopaths, who performed repeated tests.

Similar to the literature data (e.g. MOOTZ ET AL, 1989, HAWK ET AL., 1999), the intra-examiner reliability is higher than the inter-examiner reliability.

4.4. Results of the Exchange of Experiences after the Tests

The osteopaths, test persons and secretaries discussed their experiences in succession to the tests. In this discussion the following items were touched, giving hints for sources of error:

1. Different phases of concentration
While some therapists begun their test immediately after entering the room, others concentrated for a while before they started.

2. Different manual pressure
It was noticed by the test persons as well as by the secretaries, that the manual pressure applied was different. Especially O1 worked with a high pressure.

3. Different intention
One of the therapists (O13) tried - in spite of the initial introduction in the course of the investigation - to feel on the visceral level only and stated that time was too short for this technique.

4. Problems of dimension
Several osteopaths mentioned problems to project the three-dimensional information onto two dimensions.

5. Unfamiliar performance of the test

Some therapists normally do the abdominal local listening with two hands and had to face a new situation, because the test according to Barral had to be performed with only one hand.

6. Stress situation

Some osteopaths declared that they were stressed by the requirement to come to a result in the short time. Some therapists also mentioned that they felt something, but were confused because they did not know at which level.

7. Influence by the test persons

Test person P4 performed own experiments and mentioned, that every therapist sensed an O_T every time he held up his breath.

4.5. General Summary of the Results

The extent of the interexaminer reliability is very heterogeneous for different pairs of osteopaths.

In approximately half of all comparisons between the therapists, the κ-indices indicate only agreements on the level of chance. The best agreements observed feature only a fair reliability.

No significant signs for influences of tensioning caused by the lying or by nervousness of the test persons and osteopaths could be gathered but there are some indications in this direction.

The influence of the therapists' experience on the results did not turn out to be significant either, though there might be trends that the agreement increases with the experience.

The number of agreements above the level of chance is higher in the first half of the investigation than in the second half. There is evidence that the conditions have changed during the performance of the tests.

The values of the intra-examiner reliability are higher than those of the inter-examiner reliability (maximum moderate), but also here results can be observed, which are on a level of chance only.

5. Discussion

5.1. Discussion of the Method

As discussed in the first chapters, many influences onto the manual diagnostic methods may play a role, leading to errors. In this chapter restrictions and possible sources of errors will be discussed.

5.1.1. Restrictions Concerning the Therapists

The ***number of participating osteopaths,*** initially aimed at with 20, could not be reached. The stimulus for a participation on other student's studies might be too low. Thus, only 14 osteopaths took part in this investigation.

During the ***selection of osteopaths*** it was considered, that most of them were educated at the Vienna International School of Osteopathy (WSO), which should guarantee a similar state of knowledge in the field of manual diagnosis.
Nevertheless, there is a variance of experience due to the different practical activities of the individual therapists, which allows a more generalized statement about the reliability of the abdominal local listening.
The experience of the therapists was evaluated with the final year of the osteopathic education and the question, whether the osteopaths used this test "routinely", "often", "rather seldom" or "rarely".
I resigned the question of more accurate information by the therapists (estimated number/year) due to higher influences of other sources of error.
It became obvious that all therapists use this test regularly. Therefore, a classification could only be done by means of the final year of education.
This uniform answer might be caused by the non-anonymous interview in form of a list, which might have encouraged the therapists simply to repeat the first answer.

According to my experience, osteopaths have ***preferences in their diagnosis*** and therapy. In order to at least be able to recognize effects caused by this, the levels mostly worked on by the individual osteopaths (structural, visceral, fascial, cranio-sacral) were collected.
Also in this case all osteopaths stated uniformly that they had no preferences.
An anonymous questioning in advance of the study, instead of the questioning list would possibly have resulted in other outcomes.

The ***working conditions*** of the osteopaths were new to them. They had to diagnose in an unfamiliar therapy room on an unfamiliar therapy bed, which could have an influence onto the relaxation of the therapists (unrelaxed posture, inhibited breath during the manual diagnosis). For example, none of the therapists changed the height of the therapy bed in spite of different body heights.

The knowledge, that they had to come to a result in the short time of five seconds could have built up ***mental stress***, which could have been a source of distraction from sensing (Inhibited breath, thoughts during sensing).
Thus it was considered to perform a training phase, during which the test should be trained on two test persons. Since no test phase is possible in the practical work, this plan was changed and instead of this training additional evaluations of the data with and without the results on the first test persons were performed.
The result of this additional evaluation points out, that some therapists would have needed a training phase, but is not definite. The general result is not influenced significantly by omitting this point.

Additionally, the examiners themselves create a ***test situation***. One is in "competition" with other therapists and wants to sense "the right" or at least "not the wrong" quadrant. This should be avoided by the anonymity of the osteopaths and their maturity and experience, but due to psychological reasons it can not be totally precluded (egoistic consciousness, blocking thoughts during sensing), which could also be noticed in the discussion after the tests.

Prejudices in concern of the health situation of the test person, which could influence the quality of the manual diagnosis, can be caused by conscious but also unconscious acoustical, visual, olfactory and sensorial information.
These do not only comprise information from the anamnesis talk or eye-catching posture problems of the test persons, but especially unconsciously perceived impressions, like bowel sounds, smells, feeling and appearance of the skin.

By the blinding of the study (the test person is dressed except the head or covered by the cloth, respectively), and the anonymity of the test person and avoidance of talks these prejudices were excluded as far as possible.

A ***tiring of the osteopaths***, above all caused by the duration of the waiting periods, could also have occurred.
The whole procedure lasted approximately three hours, during which each therapist worked only for 1.5- 2 minutes. Tiring caused by boredom can lead to a wrong attitude, lacking concentration, wrong posture and breathing and thus to wrong results in the manual diagnosis.

The tests were held short (five seconds) on purpose, and the therapists were requested not to treat but only to test. Nevertheless, ***therapy effects***, which may occur during manual diagnosis, can not be precluded totally.
In the case that an osteopath should not keep on this request, by the randomised order of the tests it should have been possible to recognize that (under the constraint of more homogeneous results than the actual). I am aware of the fact, that many prerequisites have to be fulfilled, to recognize and preclude therapy effects efficiently.

5.1.2. Restrictions Concerning the Test Persons

The rooms were kept as quiet as possible, but due to the many changes of the osteopaths it can be assumed, that the test persons find ***relaxation*** either late or even not at all. This point was considered by the exclusion of the data of the first three osteopaths in the evaluation of the data. Hence, the influence of a accustoming phase can be worked out.

Exclusion criteria were operation scars in the abdomen and thorax, larger than the size of arthroscopical operations, because it was considered that these would shift the results in a positive way.
In order to test this theory concerning obvious symptomatic patients and in order to gain a sufficient variance of the data, test persons with small scars were included in the test program. Though, the symptoms were not evaluated.

5.1.3. Restrictions Concerning the Method

The ***comparability of the general conditions*** can not be taken as granted due to the items discussed above, which will be pronounced at each therapist and test person to a different extent. A simultaneous test by all therapists, which could at least provide the same condition of the test person for each osteopath, is not possible due to the number of osteopaths.
Thus, *temporal therapy effects* can not be precluded.

By the distinct regulation of the position, where the therapist has to put his hand, in contrary to the normal therapeutic praxis at least *local effects* can be avoided.

In spite of considering all influences on the results of manual therapeutic diagnosis, which were found in the literature, during the discussion after the ending of the investigations several points, which should have been considered additionally or more strictly, were found:

Additional to the oral explanation of the procedure, tests ***supervised*** by the conductor would have been necessary. One of the osteopaths tried to sense in the visceral field in spite of the clear instruction that the therapists should follow their first impetus and that not only visceral blockades are the theme of this test. As a side effect also therapists, who normally use both hands for the local listening could get used to the one handed test.

It is important to regulate the level, from which the information shall be gained more distinct. Much stress was caused by the fact, that some osteopaths did not know what they felt and thus were in doubt about the results. In this connection, I have to stress again, that it was my intention, to investigate the saying and reliability of this test generally and not only restricted to the visceral level.

A ***longer time for the test*** would also help to reduce stress in the osteopaths. The five seconds were chosen regarding to the preclusion of therapy effects. As it can be seen in the results of the study this restriction was too strict.

The ***pressure of the hand*** varied very much between the individual osteopaths. A regulation of the pressure could help to find a concerted level of the reason for the sensation.

One test person stated, that he could influence the results of the tests by varying his breath flow. As result, the osteopaths always found an O_T, when he held his breath. Thus, in cases like this one, as well as in other tests basing on body rhythms, a **regular breathing** of the test persons must be guaranteed. In principle, a measurement of the breathing frequency and elimination of data, gained during obvious abnormalities is possible (e.g. by means of tension test stripes). Nevertheless a regulation for the interpretation of these data must be worked out in advance.

5.2. Discussion of the Results

5.2.1. Raw Data

The variance of the results turned out to be very high. Approximately one half of the agreements between all pairs of osteopaths do not exceed the level of chance. Also in the case of the seven test persons, who had small scars in the abdomen, the quadrants where the scars were positioned, were not found predominately except in two cases.
This can either be a hint, that the scars are cleared or that in other parts of the body higher tensions prevail. Another interpretation is that the method is not sensible enough under the circumstances of this study.

5.2.2. Interexaminer Reliability

The data were evaluated under different concerns in order to be able to find out exterior influences:

1. The data of the first three tests were excluded in order to preclude influences by tensions of the test persons caused by the unfamiliar situation. That means that only the second tests of the first three osteopaths at the end of each test run were considered. 46 of all comparisons resulted in agreements between the osteopaths on basis of chance, 42 with poor reliability and three with fair reliability. The total mean of all κ-indices is 0.0048.

2. Additional to the procedure above (1), all data of the first two test persons were excluded in order to be able to visualize accustoming effects of the osteopaths. In this case 44 of all comparisons resulted in agreements between the osteopaths on basis of chance, 45 with poor reliability and two with fair reliability. The total mean of all κ-indices is 0.0047.

3. Tensioning of test persons due to the long laying was investigated by a comparison of the data of the evaluation above (1) with the data set after exclusion of the data of the last three tests (repeated tests of the first three osteopaths of each order). 48 of all comparisons resulted in agreements between the osteopaths on basis of chance, 38 with poor reliability and five with fair reliability. The total mean of all κ-indices is 0.0048.

4. Additional to the procedure above (3), all data of the first two test persons were excluded in order to be able to visualize accustoming effects of the osteopaths. In this case 47 of all comparisons resulted in agreements between the osteopaths on basis of chance, 40 with poor reliability and four with fair reliability. The total mean of all κ-indices is 0.0064.

In the comparison of the four different results gained by the methods described above, it became obvious, that the probability for agreements beyond the agreement on chance was highest, when the data of the first two test persons and the first three tests were excluded (2). Nevertheless, the degree of agreement beyond chance decreases.
The highest κ-index is 0.31, which demarks a fair reliability. It can be seen, that at least some osteopaths sense the same at the same test person.
The limitation in the course of the test, responsible for these results, have already been discussed in chapter 5.1. (Discussion of the Method), the most important factor probably is the - consciously - not fixed level of diagnosis.

The further aspects for the evaluation were:

1. the order of the tests.
2. the experience of the therapists.

By means of the calculation of κ-indices of all pairs of test positions, it was possible to find out, that the maximum agreement was reached between the 6^{th} and 7^{th} position (independent from the osteopaths) with κ=0.27.
The results of the first half of the test were compared with those of the second half by means of a subsequently performed significance test. The result shows, that the number of agreements beyond chance decreases by the proceeding of the investigation. Thus, a change of the circumstances during the study can be deduced.

An influence of the therapist's experiences might also be readable from the results. Nevertheless, statistically assessed it is not significant.
Osteopaths with longer experience in osteopathy reach agreements beyond the agreement expected on the basis of chance in 52.4% of all cases, osteopaths with shorter experience in only 38.1%

5.2.3. Intraexaminer Reliability

The intra-examiner reliability was evaluated from the data of the first three tests and the last three tests, which were performed by the same osteopaths at the same test persons. The highest agreement found was $\kappa = 0.47$, indicating a moderate reliability, one osteopath reached a fair and two examiners a poor agreement. Three of the seven comparisons resulted in agreements on the level of chance.
Due to the low number of test persons between the tests, a remembering effect can not be generally precluded for the osteopaths with high agreements.

On the other hand, low agreements might also be caused by changed conditions of the (same) test persons.

5.2.4. Additional Results

In the raw data noticeable sequences of quadrants can be observed, which are sensed more often than twice. They do not necessarily consist of the same quadrants (e.g. B/D). These sequences might be a hint, that due to a slow body rhythm different quadrants can be sensed at different times. Nevertheless, this interpretation can not be proved by the data collected in these investigations.

6. Synopsis

A universal validity of the abdominal listening could not be proved by this actual investigation, but it can be deducted from the results of this study that - in spite of the unregulated level, from which the diagnostic information should be gained - approximately half of the osteopaths achieve agreements in their diagnosis beyond the level of chance. It can be assumed that more and higher agreements can be reached, if the level is regulated more distinct.

In the actual study, as maximum kappa indices, κ=0.31, “fair reliability” was reached in the case of inter-examiner reliability, and κ=0.47, “moderate reliability” in the case of intra-examiner reliability.

The presence of adverse effects on the results, like therapy effects, accustoming and tensing problems can be supposed, but in consideration of the most limiting factor, the lack of the regulation of the diagnostic field, these effects do not affect the results significantly.

On basis of these results as next stages for the investigation of local listening can be suggested:

- Regulation of the diagnostic field.
- Preparation of the therapists in a supervised training phase.
- Regulation of the hand pressure.
- Registration of the breathing frequency of the test person and exclusion if breathing is peculiar.
- Single questionnaires for the evaluation of the therapist’s characteristics instead of a general list.
- In order to investigate cyclical body rhythms or therapy effects more osteopaths would have to take part.

7. Bibliography

Alley, J.R. (1983): The clinical value of motion palpation as a diagnostic tool: a review. J. Can. Chiro Assoc 27: 97-100.

Barral, J.-P. (2002): Lehrbuch der Viszeralen Osteopathie Band 2. München-Jena: Urban & Fischer, 5-9.

Barral, J.-P. and Mercier, P (1988): Visceral Manipulation. Seattle: Eastland Press, 21.

Beal, M.C. and Patriquin, D.A. (1995): Interexaminer agreement on palpatory diagnosis and patients self-assessment of disability: A pilot study. Journal of the American Osteopathic Association 95 (2): 97-100, 103-6.

Becker, R. (1997): Life in Motion, The osteopathic vision of Rollin E. Becker. In:

Boline P.D.; Keating J.C.; Brist, J. and Denver, G. (1988): Interexaminer reliability of palpatory evaluations of the lumbar spine. Am. J. Chiro Med 1 (1): 5-11.

Boline, P.D.; Haas, M.; Meyer, J.J.; Kassak, K.; Nelson, C. and Keating, J.C. Jr. (1993): Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality, Part II. J. Manipulative Physiol. Ther. 16 (6): 363-74.

Brooks, R.E. (Ed.): Life in Motion, The osteopathic vision of Rollin E. Becker. Portland: Rudra Press.

Burgess, P.R. and Perl, E.R.(1973): Cutaneous Mechanoreceptors and Nociceptors. In: Iggo, A.(Ed): Handbook of Sensory Physiology, vol II. New York: Springer-Verlag: 30.

Carmichael, J.P. (1987): Inter- and intraexaminer reliability of palpation for sacroiliac joint dysfunction. J. Man. Physiol Ther 10: 164-171.

Cloet, E. and Muts, R. (2005): Wie wichtig ist die visuelle Wahrnehmung in der Osteopathie. Contra - Standpunkt. DO 2005 (2): 28 - 29.

Degenhardt, B.F. et al (2005): Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. Journal of the American Osteopathic Association 105 (10): 465-473.

Dinnar, U. et al (1980) : Classsification of diagnostic tests used with osteopathic manipulation. J. Am. Osteopath. Assoc. 81: 314-321.

Dinnar, U. et al (1982) : Description of fifty diagnostic tests used with osteopathic manipulation. J. Am. Osteopath. Assoc. 79: 451-455.

Feinstein, A.R.: Clinical epidemiology: the architecture of clinical research. Philadelphia: WB Saunders: 184-6.

Fischer, R.A. and Yates, F. (1982): Statistical Tables for Biological, Agricultural and Mediacal Research 6th ed. Harlow: Longman, 146.

Fjellner, A; Bexander, C.; Faleji, R. and Strender, L.E. (1999): Interexaminer reliability in physical examination of the cervical spine. J. Manipulative Physiol Ther. 22: 511-516.

Fleiss, J.L. (1981): Statistical Methods for Rates and Proportions, 2nd ed.. New York: Wiley, 321.

Goldstein M. (1978): Discussion of clinical observations and emerging questions. In: Korr, I.M. (ed):The neurobiologic mechanisms in manipulative therapy. New York: Plenum Press: 83-98.

Haas, M. (1991): Statistical methodology for reliability studies. J. Man Physiol Ther 14: 119-132.

Haas, M. (1991): The reliability of reliability. J. Man Physiol Ther 14: 199-208.

Hawk, C.; Phongphua, C.; Bleecker, J.; Swank, L.; Lopez, D. and Rubley, T. (1999): Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. Journal of Manipulative and Physiological Therapeutics 22 (6): 382-389.

Hestbaek, L. and Leboeuf-Yde, C. (2000) (49): Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J. Man. Physiol Ther. 23: 258-275.

Hubka, M.J. and Phelan, S.P. (1994): Interexaminer reliability of palpation for cervical spine tenderness. Journal of Manipulative and Physiological Therapeutics 17 (9): 591 - 595.

Hudswell, S. et al (2005): The craniocervical flexion test using pressure biofeedback : A useful measure of cervical dysfunciton in the clinical setting. Int. Journal of Osteopathic Medicine 8 (3): 98-105.

Keating, J.C. Jr.; Bergmann, T. F.; Jacobs, G. E.; Finer, B. A. and Larson, K. (1990): Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. Journal of Manipulative and Physiological Therapeutics 13 (8): 463-470.

Kneibestol, M. Vallbo, A.B. (1970): Single unit analysis of mechanoreceptor activity from the human glubrous skin. Acta Physiol Scand 80: 178.

Koran, L.M. (1975): The reliability of clinical methods, data and judgements II. N Engl. J. Med 293: 695-701.

Kuchera, W.A. and Kappler, R.E. (2002): Musculoskeletal examination of somatic dysfunction. In: Ward, R.C. (Ed.): Foundations for Osteopathic Medicine, 2nd ed., Philadelphia: Lippincott Williams and Wilkins. 633-659.

Landis J.R. and Koch, G.G (1977): The measurement of observer agreement for categorical data. Biometrics 33: 159-174.

Lawson, A. and Calderon, L (1997): Interexaminer agreement for applied kinesiology from an interexaminer reliability study design. Perceptual and Motor Skills 84 (2): 539-546.

Ligner, B. and Van Assche, R (1994): Bildatlas der Osteopathie, Band 1. Kötzting/Bayer. Wald: Verlag für Osteopathie: 23-24.

Lockwood, M. D. and Degenhardt, B. F. (1998): Cycle-to-cycle variability attributed to the primary respiratory mechanism. Journal of the American Osteopathic Association 98 (1): 35-36, 41-43.

Lundin, M. et al. (1999): Distal pulse palpation: is it reliable?. World J. Surg. 23 (3): 252-255.

McConnell, D.G. et al. (1980): Low agreement of findings in neuromusculoskeletal examinations by a group of osteopathic physicians using their own procedures. Journal of the American Osteopathic Association 79 (7): 441-450.

Mior, S.A.; McGregor, M. and Schut, B. (1990): The role of experience in clinical accuracy. J.Man. Physiol Ther. 13: 68-71.

Mootz, R.D. et al (1989): Intra- and Interobserver Reliability of Passive Motion Palpation of the Lumbar Spine. Journal of Manipulative and Physiological Therapeutics 12: 440-445.

Moran, R. W. and Gibbons, P. (2001): Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. Journal of Manipulative and Physiological Therapeutics 24 (3): 183-190.

Muzzi, D. (2005): Palpation und Forschung. Osteopathische Medizin 2005 (1): 13-16.

Myers, K. A.; Scott, D. F.; Devine, T. J.; Johnston, A. H., Denton, M. J. and Gilfillan, I. S. (1987): Palpation of the femoral and popliteal pulses: a study of the accuracy as assessed by agreement between multiple observers. Eur J Vasc Surg 1 (4): 245-249.

Panzer, D. M. (1992): The reliability of lumbar motion palpation. J. Man. Physiol Ther 15: 518-524.

Rosner, B. (1986): Fundamentals of biostatics, 2nd ed. Boston: PWS Publishing

Ruspeckhofer, M. (2006): The influence of osteopatjhic treatment during gestation onto the course of delivery. Wien: Thesis (not yet released), WSO, 33-35

Russel, R. (1983): Diagnostic palpation of the spine: a review of procedures and assessment of their reliability. J. Man. Physiol. Ther. 6: 195-198.

Sacket, D.L.; Brian Haymes, R.; Guyatt, G.H. and Tugwell, P. (1991): Clinical Epidemiology, 2nd Ed., Philadelphia: Lippincott Williams & Williams.

Schöps, P.; Pfingsten, M. and Siebert, U. (2000): Reliabilität manualmedizinischer Untersuchungstechniken an der Halswirbelsäule. Studie zur Qualitätssicherung in der manuellen Diagnostik. Zeitschrift für Orthopädie und ihre Grenzen 138 (1): 2-7.

Smith, D. and Craige, E. (1980): Enhancement of Tactile Perception in Palpation. Circulation 62 (5): 1114-1118.

Sommerfeld, P.; Kaider, A. and Klein, P. (2004): Inter- und Intratester-Reliabilität der Palpation des "Primären Respiratorischen Mechanismus" innerhalb des "Cranialen Konzeptes". Osteopathische Medizin 2004 (4): 4-10.

Still, A.T. (1986): Philosophy and Mechanical Principles. Kirksville: Osteopathic Enterprise.

Vallbo, A.B. and Hagarth K.E. (1968): Activity form skin mechanoreceptors recorded percutaneously in awake human subjects. Exp Neurol. 21: 270.

Weber, K.-H. (2000): The Importance of Sensing and its relationship to the philosophy of Osteopathy. Vienna: Thesis (not yet released). WSO.

Woodword, M. (1999): Epidemiology. Study design and data analysis. Boca Rata-London-New York: Chapman and Hall.

8. Table of Illustrations

Annex 1

Experience of the Osteopaths

Osteopath (ID)	Final year
O1	2004
O2	2005
O3	1998
O4	2000
O5	2001
O6	2005
O7	2005
O8	1998
O9	1999
O10	2005
O11	2005
O12	2002
O13	2000
O14	2001

Annex 2

Statistical Data Sheet

Patient (ID): P

Date:

Scar: no / yes

in Quadrant: (A, B, C, D, 0)

Position	Osteopath (ID)	Quadrant (A, B, C, D, 0)
1	O	
2	O	
3	O	
4	O	
5	O	
6	O	
7	O	
8	O	
9	O	
10	O	
11	O	
12	O	
13	O	
14	O	
15	O	
16	O	
17	O	
18	O	
19	O	
20	O	
= 1	O	
= 2	O	
= 3	O	

Annex 3

Results

n... number of observations
Ex1... Osteopath 1 (ID)
Ex2... Osteopath 2 (ID)
Kappa... Cohen's Kappa
95%l... lower 95%-confidence interval
95%u... upper 95%-confidence interval
Sign... Significance of the alternative-hypothesis,
that the result is not on the basis of chance

Annex 3a

Results

(Positions 4-17)

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	O1	O2	-0.12	-0.33	0.08	n.s.
15	O1	O3	0.03	-0.11	0.17	n.s.
15	O1	O4	0.03	-0.18	0.25	n.s.
15	O1	O5	-0.09	-0.28	0.10	n.s.
15	O1	O6	0.06	-0.13	0.26	n.s.
15	O1	O7	-0.13	-0.37	0.12	n.s.
15	O1	O8	-0.02	-0.20	0.17	n.s.
15	O1	O9	0.02	-0.20	0.24	n.s.
15	O1	O10	0.03	-0.19	0.24	n.s.
15	O1	O11	-0.16	-0.40	0.08	n.s.
15	O1	O12	-0.13	-0.37	0.11	n.s.
15	O1	O13	-0.04	-0.20	0.11	n.s.
15	O1	O14	-0.05	-0.25	0.15	n.s.
15	O2	O1	-0.12	-0.33	0.08	n.s.
15	O2	O3	-0.23	-0.50	0.05	n.s.
15	O2	O4	0.04	-0.11	0.19	n.s.
15	O2	O5	-0.10	-0.26	0.06	n.s.
15	O2	O6	-0.07	-0.29	0.14	n.s.
15	O2	O7	0.08	-0.09	0.25	n.s.
15	O2	O8	-0.05	-0.30	0.20	n.s.
15	O2	O9	-0.05	-0.26	0.15	n.s.
15	O2	O10	-0.07	-0.24	0.09	n.s.
15	O2	O11	0.06	-0.13	0.25	n.s.
15	O2	O12	0.10	-0.11	0.31	n.s.
15	O2	O13	0.17	-0.02	0.36	n.s.
15	O2	O14	-0.03	-0.11	0.04	n.s.
15	O3	O1	0.03	-0.11	0.17	n.s.
15	O3	O2	-0.23	-0.50	0.05	n.s.
15	O3	O4	-0.07	-0.19	0.04	n.s.
15	O3	O5	0.18	0.00	0.36	n.s.
15	O3	O6	0.06	-0.13	0.24	n.s.
15	O3	O7	-0.02	-0.15	0.10	n.s.
15	O3	O8	0.14	-0.09	0.37	n.s.
15	O3	O9	0.05	-0.12	0.22	n.s.
15	O3	O10	-0.05	-0.19	0.09	n.s.
15	O3	O11	-0.06	-0.21	0.10	n.s.
15	O3	O12	0.03	-0.10	0.17	n.s.
15	O3	O13	-0.04	-0.26	0.18	n.s.
15	O3	O14	-0.03	-0.10	0.03	n.s.
15	O4	O1	0.03	-0.18	0.25	n.s.
15	O4	O2	0.04	-0.11	0.19	n.s.
15	O4	O3	-0.07	-0.19	0.04	n.s.
15	O4	O5	-0.06	-0.22	0.11	n.s.
15	O4	O6	-0.03	-0.19	0.12	n.s.
15	O4	O7	-0.08	-0.29	0.13	n.s.
15	O4	O8	-0.02	-0.16	0.12	n.s.
15	O4	O9	-0.17	-0.35	0.02	n.s.
15	O4	O10	0.15	-0.05	0.34	n.s.

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	O4	O11	-0.09	-0.29	0.10	n.s.
15	O4	O12	0.09	-0.10	0.29	n.s.
15	O4	O13	-0.09	-0.23	0.04	n.s.
15	O4	O14	-0.21	-0.41	-0.01	p < 0.05
15	O5	O1	-0.09	-0.28	0.10	n.s.
15	O5	O2	-0.10	-0.26	0.06	n.s.
15	O5	O3	0.18	0.00	0.36	n.s.
15	O5	O4	-0.06	-0.22	0.11	n.s.
15	O5	O6	0.11	-0.05	0.26	n.s.
15	O5	O7	0.12	-0.08	0.32	n.s.
15	O5	O8	-0.03	-0.22	0.16	n.s.
15	O5	O9	0.06	-0.12	0.25	n.s.
15	O5	O10	0.07	-0.12	0.26	n.s.
15	O5	O11	-0.04	-0.25	0.16	n.s.
15	O5	O12	0.03	-0.10	0.17	n.s.
15	O5	O13	-0.15	-0.36	0.05	n.s.
15	O5	O14	-0.01	-0.21	0.20	n.s.
15	O6	O1	0.06	-0.13	0.26	n.s.
15	O6	O2	-0.07	-0.29	0.14	n.s.
15	O6	O3	0.06	-0.13	0.24	n.s.
15	O6	O4	-0.03	-0.19	0.12	n.s.
15	O6	O5	0.11	-0.05	0.26	n.s.
15	O6	O7	0.03	-0.13	0.19	n.s.
15	O6	O8	-0.03	-0.24	0.18	n.s.
15	O6	O9	0.31	0.14	0.49	p < 0.001
15	O6	O10	0.10	-0.07	0.27	n.s.
15	O6	O11	-0.08	-0.26	0.11	n.s.
15	O6	O12	-0.09	-0.27	0.09	n.s.
15	O6	O13	0.03	-0.11	0.18	n.s.
15	O6	O14	-0.08	-0.20	0.04	n.s.
15	O7	O1	-0.13	-0.37	0.12	n.s.
15	O7	O2	0.08	-0.09	0.25	n.s.
15	O7	O3	-0.02	-0.15	0.10	n.s.
15	O7	O4	-0.08	-0.29	0.13	n.s.
15	O7	O5	0.12	-0.08	0.32	n.s.
15	O7	O6	0.03	-0.13	0.19	n.s.
15	O7	O8	0.12	-0.03	0.26	n.s.
15	O7	O9	-0.10	-0.31	0.12	n.s.
15	O7	O10	-0.05	-0.25	0.15	n.s.
15	O7	O11	0.04	-0.17	0.24	n.s.
15	O7	O12	0.21	-0.06	0.47	n.s.
15	O7	O13	0.07	-0.09	0.24	n.s.
15	O7	O14	0.20	-0.03	0.43	n.s.
15	O8	O1	-0.02	-0.20	0.17	n.s.
15	O8	O2	-0.05	-0.30	0.20	n.s.
15	O8	O3	0.14	-0.09	0.37	n.s.
15	O8	O4	-0.02	-0.16	0.12	n.s.
15	O8	O5	-0.03	-0.22	0.16	n.s.

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	O8	O6	-0.03	-0.24	0.18	n.s.
15	O8	O7	0.12	-0.03	0.26	n.s.
15	O8	O9	0.05	-0.14	0.25	n.s.
15	O8	O10	-0.01	-0.18	0.17	n.s.
15	O8	O11	0.07	-0.12	0.26	n.s.
15	O8	O12	0.11	-0.04	0.26	n.s.
15	O8	O13	0.02	-0.19	0.22	n.s.
15	O8	O14	0.02	-0.06	0.11	n.s.
15	O9	O1	0.02	-0.20	0.24	n.s.
15	O9	O2	-0.05	-0.26	0.15	n.s.
15	O9	O3	0.05	-0.12	0.22	n.s.
15	O9	O4	-0.17	-0.35	0.02	n.s.
15	O9	O5	0.06	-0.12	0.25	n.s.
15	O9	O6	0.31	0.14	0.49	p < 0.001
15	O9	O7	-0.10	-0.31	0.12	n.s.
15	O9	O8	0.05	-0.14	0.25	n.s.
15	O9	O10	0.11	-0.11	0.33	n.s.
15	O9	O11	0.00	-0.18	0.18	n.s.
15	O9	O12	0.08	-0.13	0.29	n.s.
15	O9	O13	0.04	-0.16	0.23	n.s.
15	O9	O14	-0.04	-0.17	0.09	n.s.
15	O10	O1	0.03	-0.19	0.24	n.s.
15	O10	O2	-0.07	-0.24	0.09	n.s.
15	O10	O3	-0.05	-0.19	0.09	n.s.
15	O10	O4	0.15	-0.05	0.34	n.s.
15	O10	O5	0.07	-0.12	0.26	n.s.
15	O10	O6	0.10	-0.07	0.27	n.s.
15	O10	O7	-0.05	-0.25	0.15	n.s.
15	O10	O8	-0.01	-0.18	0.17	n.s.
15	O10	O9	0.11	-0.11	0.33	n.s.
15	O10	O11	0.15	-0.04	0.33	n.s.
15	O10	O12	-0.03	-0.20	0.15	n.s.
15	O10	O13	-0.08	-0.25	0.10	n.s.
15	O10	O14	0.01	-0.15	0.17	n.s.
15	O11	O1	-0.16	-0.40	0.08	n.s.
15	O11	O2	0.06	-0.13	0.25	n.s.
15	O11	O3	-0.06	-0.21	0.10	n.s.
15	O11	O4	-0.09	-0.29	0.10	n.s.
15	O11	O5	-0.04	-0.25	0.16	n.s.
15	O11	O6	-0.08	-0.26	0.11	n.s.
15	O11	O7	0.04	-0.17	0.24	n.s.
15	O11	O8	0.07	-0.12	0.26	n.s.
15	O11	O9	0.00	-0.18	0.18	n.s.
15	O11	O10	0.15	-0.04	0.33	n.s.
15	O11	O12	0.06	-0.13	0.24	n.s.
15	O11	O13	-0.03	-0.18	0.11	n.s.
15	O11	O14	0.10	-0.10	0.30	n.s.
15	O12	O1	-0.13	-0.37	0.11	n.s.

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	O12	O2	0.10	-0.11	0.31	n.s.
15	O12	O3	0.03	-0.10	0.17	n.s.
15	O12	O4	0.09	-0.10	0.29	n.s.
15	O12	O5	0.03	-0.10	0.17	n.s.
15	O12	O6	-0.09	-0.27	0.09	n.s.
15	O12	O7	0.21	-0.06	0.47	n.s.
15	O12	O8	0.11	-0.04	0.26	n.s.
15	O12	O9	0.08	-0.13	0.29	n.s.
15	O12	O10	-0.03	-0.20	0.15	n.s.
15	O12	O11	0.06	-0.13	0.24	n.s.
15	O12	O13	0.02	-0.12	0.16	n.s.
15	O12	O14	0.02	-0.14	0.18	n.s.
15	O13	O1	-0.04	-0.20	0.11	n.s.
15	O13	O2	0.17	-0.02	0.36	n.s.
15	O13	O3	-0.04	-0.26	0.18	n.s.
15	O13	O4	-0.09	-0.23	0.04	n.s.
15	O13	O5	-0.15	-0.36	0.05	n.s.
15	O13	O6	0.03	-0.11	0.18	n.s.
15	O13	O7	0.07	-0.09	0.24	n.s.
15	O13	O8	0.02	-0.19	0.22	n.s.
15	O13	O9	0.04	-0.16	0.23	n.s.
15	O13	O10	-0.08	-0.25	0.10	n.s.
15	O13	O11	-0.03	-0.18	0.11	n.s.
15	O13	O12	0.02	-0.12	0.16	n.s.
15	O13	O14	-0.01	-0.12	0.11	n.s.
15	O14	O1	-0.05	-0.25	0.15	n.s.
15	O14	O2	-0.03	-0.11	0.04	n.s.
15	O14	O3	-0.03	-0.10	0.03	n.s.
15	O14	O4	-0.21	-0.41	-0.01	$p < 0.05$
15	O14	O5	-0.01	-0.21	0.20	n.s.
15	O14	O6	-0.08	-0.20	0.04	n.s.
15	O14	O7	0.20	-0.03	0.43	n.s.
15	O14	O8	0.02	-0.06	0.11	n.s.
15	O14	O9	-0.04	-0.17	0.09	n.s.
15	O14	O10	0.01	-0.15	0.17	n.s.
15	O14	O11	0.10	-0.10	0.30	n.s.
15	O14	O12	0.02	-0.14	0.18	n.s.
15	O14	O13	-0.01	-0.12	0.11	n.s.

Annex 3b

Results

(Positions 1-14)

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	1	2	-0.12	-0.32	0.09	n.s.
15	1	3	-0.08	-0.26	0.10	n.s.
15	1	4	0.04	-0.15	0.23	n.s.
15	1	5	-0.07	-0.24	0.10	n.s.
15	1	6	-0.01	-0.19	0.18	n.s.
15	1	7	-0.17	-0.43	0.09	n.s.
15	1	8	0.00	-0.18	0.18	n.s.
15	1	9	-0.07	-0.29	0.16	n.s.
15	1	10	-0.13	-0.34	0.08	n.s.
15	1	11	-0.06	-0.29	0.17	n.s.
15	1	12	-0.26	-0.51	0.00	p < 0.05
15	1	13	-0.04	-0.20	0.11	n.s.
15	1	14	-0.05	-0.26	0.15	n.s.
15	2	1	-0.12	-0.32	0.09	n.s.
15	2	3	-0.12	-0.36	0.12	n.s.
15	2	4	0.00	-0.12	0.12	n.s.
15	2	5	-0.09	-0.25	0.06	n.s.
15	2	6	-0.07	-0.29	0.14	n.s.
15	2	7	0.08	-0.09	0.25	n.s.
15	2	8	-0.05	-0.30	0.20	n.s.
15	2	9	-0.05	-0.26	0.15	n.s.
15	2	10	-0.07	-0.24	0.09	n.s.
15	2	11	0.06	-0.12	0.25	n.s.
15	2	12	0.10	-0.11	0.31	n.s.
15	2	13	0.07	-0.10	0.23	n.s.
15	2	14	-0.04	-0.12	0.05	n.s.
15	3	1	-0.08	-0.26	0.10	n.s.
15	3	2	-0.12	-0.36	0.12	n.s.
15	3	4	-0.11	-0.25	0.03	n.s.
15	3	5	0.01	-0.15	0.17	n.s.
15	3	6	-0.08	-0.27	0.10	n.s.
15	3	7	-0.03	-0.21	0.15	n.s.
15	3	8	0.04	-0.16	0.24	n.s.
15	3	9	0.05	-0.16	0.25	n.s.
15	3	10	0.03	-0.14	0.19	n.s.
15	3	11	-0.06	-0.22	0.11	n.s.
15	3	12	0.25	0.04	0.46	p < 0.025
15	3	13	0.21	0.02	0.41	p < 0.05
15	3	14	0.08	-0.01	0.18	n.s.
15	4	1	0.04	-0.15	0.23	n.s.
15	4	2	0.00	-0.12	0.12	n.s.
15	4	3	-0.11	-0.25	0.03	n.s.
15	4	5	0.01	-0.09	0.12	n.s.
15	4	6	-0.01	-0.15	0.12	n.s.
15	4	7	-0.01	-0.20	0.18	n.s.
15	4	8	0.02	-0.07	0.12	n.s.
15	4	9	-0.15	-0.32	0.02	n.s.
15	4	10	0.23	0.06	0.41	p < 0.01

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	4	11	-0.06	-0.23	0.11	n.s.
15	4	12	-0.02	-0.21	0.18	n.s.
15	4	13	-0.05	-0.21	0.11	n.s.
15	4	14	-0.03	-0.23	0.16	n.s.
15	5	1	-0.07	-0.24	0.10	n.s.
15	5	2	-0.09	-0.25	0.06	n.s.
15	5	3	0.01	-0.15	0.17	n.s.
15	5	4	0.01	-0.09	0.12	n.s.
15	5	6	0.03	-0.12	0.19	n.s.
15	5	7	0.08	-0.09	0.25	n.s.
15	5	8	-0.04	-0.24	0.16	n.s.
15	5	9	0.08	-0.09	0.25	n.s.
15	5	10	0.08	-0.10	0.26	n.s.
15	5	11	0.07	-0.12	0.25	n.s.
15	5	12	0.07	-0.04	0.18	n.s.
15	5	13	-0.03	-0.23	0.16	n.s.
15	5	14	0.05	-0.12	0.22	n.s.
15	6	1	-0.01	-0.19	0.18	n.s.
15	6	2	-0.07	-0.29	0.14	n.s.
15	6	3	-0.08	-0.27	0.10	n.s.
15	6	4	-0.01	-0.15	0.12	n.s.
15	6	5	0.03	-0.12	0.19	n.s.
15	6	7	0.03	-0.13	0.19	n.s.
15	6	8	-0.03	-0.24	0.18	n.s.
15	6	9	0.31	0.14	0.49	p < 0.001
15	6	10	0.10	-0.07	0.27	n.s.
15	6	11	-0.07	-0.26	0.11	n.s.
15	6	12	-0.09	-0.27	0.09	n.s.
15	6	13	-0.01	-0.15	0.12	n.s.
15	6	14	-0.09	-0.23	0.04	n.s.
15	7	1	-0.17	-0.43	0.09	n.s.
15	7	2	0.08	-0.09	0.25	n.s.
15	7	3	-0.03	-0.21	0.15	n.s.
15	7	4	-0.01	-0.20	0.18	n.s.
15	7	5	0.08	-0.09	0.25	n.s.
15	7	6	0.03	-0.13	0.19	n.s.
15	7	8	0.12	-0.03	0.26	n.s.
15	7	9	-0.10	-0.31	0.12	n.s.
15	7	10	-0.05	-0.25	0.15	n.s.
15	7	11	0.02	-0.20	0.23	n.s.
15	7	12	0.21	-0.06	0.47	n.s.
15	7	13	0.08	-0.08	0.24	n.s.
15	7	14	-0.02	-0.23	0.20	n.s.
15	8	1	0.00	-0.18	0.18	n.s.
15	8	2	-0.05	-0.30	0.20	n.s.
15	8	3	0.04	-0.16	0.24	n.s.
15	8	4	0.02	-0.07	0.12	n.s.
15	8	5	-0.04	-0.24	0.16	n.s.

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	8	6	-0.03	-0.24	0.18	n.s.
15	8	7	0.12	-0.03	0.26	n.s.
15	8	9	0.05	-0.14	0.25	n.s.
15	8	10	-0.01	-0.18	0.17	n.s.
15	8	11	0.01	-0.17	0.18	n.s.
15	8	12	0.11	-0.04	0.26	n.s.
15	8	13	0.01	-0.16	0.17	n.s.
15	8	14	0.01	-0.09	0.11	n.s.
15	9	1	-0.07	-0.29	0.16	n.s.
15	9	2	-0.05	-0.26	0.15	n.s.
15	9	3	0.05	-0.16	0.25	n.s.
15	9	4	-0.15	-0.32	0.02	n.s.
15	9	5	0.08	-0.09	0.25	n.s.
15	9	6	0.31	0.14	0.49	p < 0.001
15	9	7	-0.10	-0.31	0.12	n.s.
15	9	8	0.05	-0.14	0.25	n.s.
15	9	10	0.11	-0.11	0.33	n.s.
15	9	11	-0.01	-0.20	0.18	n.s.
15	9	12	0.08	-0.13	0.29	n.s.
15	9	13	0.07	-0.11	0.25	n.s.
15	9	14	-0.03	-0.17	0.10	n.s.
15	10	1	-0.13	-0.34	0.08	n.s.
15	10	2	-0.07	-0.24	0.09	n.s.
15	10	3	0.03	-0.14	0.19	n.s.
15	10	4	0.23	0.06	0.41	p < 0.01
15	10	5	0.08	-0.10	0.26	n.s.
15	10	6	0.10	-0.07	0.27	n.s.
15	10	7	-0.05	-0.25	0.15	n.s.
15	10	8	-0.01	-0.18	0.17	n.s.
15	10	9	0.11	-0.11	0.33	n.s.
15	10	11	0.07	-0.12	0.26	n.s.
15	10	12	-0.03	-0.20	0.15	n.s.
15	10	13	-0.06	-0.22	0.11	n.s.
15	10	14	0.01	-0.17	0.18	n.s.
15	11	1	-0.06	-0.29	0.17	n.s.
15	11	2	0.06	-0.12	0.25	n.s.
15	11	3	-0.06	-0.22	0.11	n.s.
15	11	4	-0.06	-0.23	0.11	n.s.
15	11	5	0.07	-0.12	0.25	n.s.
15	11	6	-0.07	-0.26	0.11	n.s.
15	11	7	0.02	-0.20	0.23	n.s.
15	11	8	0.01	-0.17	0.18	n.s.
15	11	9	-0.01	-0.20	0.18	n.s.
15	11	10	0.07	-0.12	0.26	n.s.
15	11	12	0.03	-0.17	0.23	n.s.
15	11	13	-0.03	-0.19	0.12	n.s.
15	11	14	-0.03	-0.23	0.16	n.s.
15	12	1	-0.26	-0.51	0.00	p < 0.05

n	Ex1	Ex2	Kappa	95%l	95%u	Sign.
15	12	2	0.10	-0.11	0.31	n.s.
15	12	3	0.25	0.04	0.46	p < 0.025
15	12	4	-0.02	-0.21	0.18	n.s.
15	12	5	0.07	-0.04	0.18	n.s.
15	12	6	-0.09	-0.27	0.09	n.s.
15	12	7	0.21	-0.06	0.47	n.s.
15	12	8	0.11	-0.04	0.26	n.s.
15	12	9	0.08	-0.13	0.29	n.s.
15	12	10	-0.03	-0.20	0.15	n.s.
15	12	11	0.03	-0.17	0.23	n.s.
15	12	13	0.10	-0.03	0.24	n.s.
15	12	14	0.05	-0.11	0.20	n.s.
15	13	1	-0.04	-0.20	0.11	n.s.
15	13	2	0.07	-0.10	0.23	n.s.
15	13	3	0.21	0.02	0.41	p < 0.05
15	13	4	-0.05	-0.21	0.11	n.s.
15	13	5	-0.03	-0.23	0.16	n.s.
15	13	6	-0.01	-0.15	0.12	n.s.
15	13	7	0.08	-0.08	0.24	n.s.
15	13	8	0.01	-0.16	0.17	n.s.
15	13	9	0.07	-0.11	0.25	n.s.
15	13	10	-0.06	-0.22	0.11	n.s.
15	13	11	-0.03	-0.19	0.12	n.s.
15	13	12	0.10	-0.03	0.24	n.s.
15	13	14	0.05	-0.08	0.18	n.s.
15	14	1	-0.05	-0.26	0.15	n.s.
15	14	2	-0.04	-0.12	0.05	n.s.
15	14	3	0.08	-0.01	0.18	n.s.
15	14	4	-0.03	-0.23	0.16	n.s.
15	14	5	0.05	-0.12	0.22	n.s.
15	14	6	-0.09	-0.23	0.04	n.s.
15	14	7	-0.02	-0.23	0.20	n.s.
15	14	8	0.01	-0.09	0.11	n.s.
15	14	9	-0.03	-0.17	0.10	n.s.
15	14	10	0.01	-0.17	0.18	n.s.
15	14	11	-0.03	-0.23	0.16	n.s.
15	14	12	0.05	-0.11	0.20	n.s.
15	14	13	0.05	-0.08	0.18	n.s.