



RELIABILITY OF THE OBJECTIVE SYNCHRONOUS TEST AS USED IN ADVANCED BIOSTRUCTURAL CORRECTION™ TO ASSESS FOR L5 DYSFUNCTION (ROOSTA-L5)

Beau Woods, BSc(Chiro), BChiro, MPH¹, Neil Thomas, DC, MChiro, PgD (AdvPrac), PgC (MedEd), PgC (MedUS), PgC (ProfDev), AFHEA, FEAC, FRCC (Ortho), Doctoral-(cand)², Melinda Stanners BHSci, PhD(Medicine)³, Kelly Holt, BSc, BChiro, PGDipHSc, PhD⁴

ABSTRACT

Objective: To investigate the intra-examiner and inter-examiner reliability of the 'Objective Synchronous Test' (OST) as used in Advanced BioStructural Correction™ (ABC™) to assess for dysfunction at the 5th lumbar vertebrae (L5) in participants with and without spine pain, using both experienced and inexperienced ABC™ practitioners.

Methods: Four examiners (2 basic level, 2 instructor level) examined a sample of symptomatic (n = 7) and asymptomatic (n = 16) participants using the OST, which is used as a primary indicator for spinal manipulation of L5 as part of the ABC™ protocol. L5 was tested and retested using the OST as an indicator for correction of L5. The OST was performed on every participant (n=23) by all 4 examiners. Intra-examiner and inter-examiner reliability was reported with observed agreement (Po) and analyzed with kappa (k).

Results: Observed agreement for intra-examiner reliability ranged from 56.5% to 95.7% and averaged 71.8%. Estimates of intra-examiner reliability for L5 ranged from $k = 0.13$ to 0.91 and were higher than estimates of inter-examiner reliability. Inter-examiner reliability estimates for L5 ranged from $k = 0.42$ to 0.47 . The median observed agreement between paired examiners for L5 ranged from 65.9 to 73.9%.

Conclusion: In this study, the inter-examiner reliability of the OST was generally found to be moderate when assessing for L5 dysfunction. The intra-examiner reliability shown for more experienced (instructor level) examiners was higher than less experienced (basic level) examiners. (*J Contemporary Chiropr 2023;6:1-7*)

Key Indexing Terms: Chiropractic; Observer Variation; Reproducibility of Results; Low Back Pain; Lumbar Vertebrae;

INTRODUCTION

The identification of the site of spinal manipulation is important to the practice of chiropractic, yet is fraught, due to the lack of a 'gold standard' for identifying the 'manipulable lesion' which is universally agreed upon, as discussed in the systematic review by Triano. (1) Treatment of dysfunction of the neuromusculoskeletal system includes manual manipulation procedures directed toward normalizing alterations of the locomotor

¹ Australian Chiropractic College, Adelaide, SA, Australia

² Teesside University, Middlesborough, England,

³ Torrens University, Adelaide, SA, Australia,

⁴ New Zealand College of Chiropractic. Mt. Wellington, Auckland, New Zealand

system (which includes the spine). (1) The wide range of techniques used to determine the site for applying spinal manipulation have been loosely aggregated into related categories; pain, asymmetry, relative range of motion, changes in tissue temperature/texture/tone, and findings from special tests (PARTS). (1) One of the inclusion criteria for the aforementioned Triano review was that 'at least some of the subjects must have been symptomatic or have had a known anatomical anomaly.' (1)

In general, the stronger and more favorable evidence is for those procedures that take a direct measure of the presumptive site of care, such as methods involving pain provocation upon palpation or localized tissue examination (1). Assessing commonly used procedures for reliability is thus a priority for practitioners and researchers alike in order to improve the quality of care that is provided. (1- 3)

ABC™ is a manual therapy method that is practiced predominantly by chiropractors in Australia but also by other healthcare professionals internationally. In Australia, the association representing ABC practitioners, Advanced Biostructural Correction Australasia (ABCA), reports 94 Australian members, which equates to 1.5% of the 6,147 chiropractors currently registered in Australia. (4) The determination of what body structures require intervention as part of the ABC™ protocol is largely dependent on the findings of the objective synchronous test (OST) along with other elements of clinical decision-making. The OST has never previously been evaluated for reliability. The OST is performed while the participant is standing with the examiner standing behind. The examiner would lightly push the spinous of L5 in an anterior direction and then immediately perform the OST, as described below.

The ABC™ protocol employs a number of elements: 'meningeal releases,' that are long-lever manipulations aimed at decreasing 'dural tension,' and specific adjustments or corrections to the spine, pelvis, ribs, feet and legs. (5)

Low back pain is the leading worldwide cause of years lost to disability and its burden is growing. (6) L5 is commonly identified as a symptomatic area in chiropractic clinical practice and manual therapy due to the unique properties of this part of the spine such as its kinematic instability and propensity for joint/ disc degeneration when compared to upper lumbar vertebrae. (2,3,7,8) Previous studies have measured intra-examiner and inter-examiner reliability of other

manual therapy testing methods for L5, such as static palpation, motion palpation and elicited tenderness, with k values ranging from less than chance to almost perfect. (2,9-11) Anatomical identification of the L5 spinous process has shown to be problematic, with static identification by trained physiotherapists as low as 45% when compared to a 'gold-standard' radiographic analysis. (12) The scope of this paper is focused on L5 with future research to widen and include evaluation of other components of the ABC™ method.

A secondary aim of this study was to compare the reliability estimates of chiropractors with extensive experience practicing ABC™ versus chiropractors with lesser experience. A common criticism of studies in this area is that inexperienced students are commonly used as examiners. (1)

METHODS

Participants

Participants were recruited from the general public via social media and word of mouth invitation in Melbourne, Australia. Potential participants were directed to a webpage created via callforparticipants.com and were provided with an information sheet via email as well as given a hard copy in person that outlined details of the study. Participants were provided a \$30 shopping voucher to reimburse travel costs and time, and all gave informed consent. Eligibility criteria required participants to be aged between 18 and 80 regardless of their spine pain status. Subjects were excluded if they had any history or physical signs of serious spinal pathology or spinal nerve root problems. Exclusion criteria also applied to those with visibly identifiable characteristics such as tattoos or skin lesions which could potentially cue the examiners, or those who experience any discomfort while standing, have exhibited obvious pain behaviors particularly with movement, or who have a body mass index (BMI) ≥ 30 kg/m²

Examiners

Four examiners (raters) were recruited from the clinical teaching faculty of Advanced BioStructural Correction™ Australasia and the pool of ABC™ practitioners located in Melbourne, Australia.

Identification of Anatomical Structure

The lead investigator marked the exposed skin of the participants with a non-toxic black marker pen over the

spinous process of L5. The skin was marked because previous research has shown that the accuracy of surface anatomy identification can be poor and this method would avoid variation in the identification of the L5 spinous process. (13)

Examiner Training

It was assumed that all examiners would not need additional training prior to the study as the testing procedures and OST are a standard part of ABC™ practitioner training and practice. Examiners received verbal instruction on how the study would be carried out in regards to logistics and protocols to ensure blinding. The following testing procedures were employed.

Objective Synchronous Test

The OST is performed by the following procedure as developed and described by Jutkowitz. (5) With the participant in the standing position facing away from the examiner, the examiner stands behind the participant. The examiner tests L5 by pushing lightly on its spinous process 5mm in a posterior to anterior direction. The OST would then immediately be performed. The OST is performed with both hands made into a fist and the thumbs held in extension, the tips of the thumbs are placed bilaterally 5mm lateral to the external occipital protuberance (EOP). The left thumb is slid inferiorly off the EOP into the indentation made over the arch of C1. While the left thumb is kept in place, the right thumb now follows the same action on the right. The level of the thumbs are then compared giving the examiner a finding of either thumb down on the left, thumb down on the right or even. In practice, this is compared to the known 'breakdown side' of the patient and correction of L5 is performed if indicated by a positive finding, i.e., the thumb is down on the side of breakdown. (5)

Experimental Procedure

The study used a repeated measures design on a single day to investigate intra-examiner and inter-examiner reliability, as per Kmita. (14) Each subject completed a standardized case history form prior to enrolment in the study. Ethics approval for this study was provided by Torrens University Australia Human Research Ethics Committee, protocol number CRM:0005289.

Stage 1: groups of 4 participants at a time changed into medical gowns and were allocated to a corner of the room standing facing outwards towards the wall. Participants were instructed to remain quiet, avoid moving their body, not to talk and examiners were to

avoid looking into the face of the participant. Examiners would mark data collection forms with their allocated number and the number of their participant. The participant's allocated number was marked on their hand in non-toxic marker. Examiners would test each participant and then place the completed data collection form face down in front of the investigator. When signaled, all participants rotated to the next examiner in a clockwise direction. After testing all participants, examiners were given a 5-minute break and then the study procedure was repeated. As participants arrived in waves, there was always another group available between the test and re-test groups.

Stage 2: the same process was followed except that participants were randomly allocated to a new starting position and the data collection forms were marked 're-test.'

These methods were used to aid examiner blinding to the identity of the participants and covered all other aspects of the participant's body that may have acted as a cue to recall prior examination findings. The chief investigator controlled the flow of the examiners and ensured that all participants and examiners complied with blinding protocols. For the full duration of the study, examiners were blinded to other examiners findings and also to their own prior findings. Between the assessment of each participant, examiners were expected to 'reset' the participant by gently shaking their shoulders. This is a requirement with ABC™ to prevent a false OST reading (5).

Data Analysis

The raw data were tabulated and transcribed into contingency tables within a data analysis and statistical software package (IBM SPSS 29 - 2022). Percentage agreement (Po), Cohen's kappa (k) and Fleiss' kappa (k) were then calculated in using SPSS. For each examiner, test-re-test data were used to estimate the intra-examiner reliability as described by Cohen. (15) The inter-examiner reliability of each pair of examiners was also calculated according to Cohen. (15) The reliability of all examiners was calculated as described by Fleiss, as reliability of more than 2 examiners cannot be calculated using Cohens method. (16)

Kappa is the most commonly used statistic to evaluate the reliability of categorical tests and estimates the level of agreement between 2 measurements in excess of chance agreement (15). Kappa ranges from 1 to -1. A kappa of 0 represents agreement that is equal to chance alone, whereas positive values of kappa indicate

Table 1. Interpretation of Kappa Coefficient - According to Landis and Koch (17)

| VALUE OF KAPPA | AGREEMENT |
|----------------|------------------|
| -1 to zero | Less than chance |
| Zero | Chance alone |
| 0.00 - 0.20 | Slight |
| 0.21 - 0.40 | Fair |
| 0.41-0.60 | Moderate |
| 0.61-0.80 | Substantial |
| 0.81-0.99 | Almost perfect |

agreement beyond that of chance. Negative kappa values indicate that chance agreement was greater than observed agreement. (17).] Landis and Koch published a qualitative scale for kappa, in which the magnitude of kappa was proposed to indicate the level of agreement achieved (Table 1). (17) There is disagreement regarding the use of these qualitative indicators, predominantly because the prevalence of the sign being measured in the sample can influence kappa and lead to erratic estimates. (15) In addition, there is no evidence to suggest that the criteria suggested by Landis and Koch are valid. In the field of physical medicine, some authors will accept a kappa of > 0.40 as clinically relevant, while others require kappa to be > 0.6. (18)

With 4 examiners and 23 participants, the study has approximately 80% power to detect kappa statistics of approximately 0.40 with a 5% significance level. (19) Expected agreement is chance alone, or a kappa value of 0.

RESULTS

A total of 23 participants were recruited for this study (n = 23). There were no drop-outs. Of the 23 participants that comprised the final sample, 16 had no history of spinal pain and 7 had current or recent non-specific spinal pain. The characteristics of the participants are detailed in Table 2. Four chiropractors with different levels of clinical experience (20,12, 3 & 1.5 years) and who all work in private practice were included as examiners (raters). The 2 most experienced examiners are also instructors in ABC™. The 2 less experienced ABC™ practitioners were certified at basic level (Level 1). No additional training was undertaken as it was assumed that all examiners were sufficiently experienced to participate in the study. Extensive examiner training

Table 2. Characteristics of the Study Subjects

| PARTICIPANTS (N = 23) | CHARACTERISTICS |
|--------------------------------|--|
| Age range | 18-42 |
| Gender | 12F : 11M |
| Spinal Pain (any region) | n=7, central and bilateral pain distribution; not below the gluteal region |
| Neck Pain | n=2 |
| Thoracic Pain | n=0 |
| Lumbo-pelvic pain | n=5 |
| History of spine pain duration | 2-240 (range in months) |
| Height | Symptomatic 161-193cm Asymptomatic 155-175 cm |
| Weight | Symptomatic 65-92 kg Asymptomatic 55-80 kg |

specifically for the purpose of increasing examiner agreement is likely to improve reliability. (20)

Intra-examiner Reliability

The intra-examiner reliability for testing of L5 was calculated from test-re-test data and is presented in table 3 (percentage agreement) and table 4 (intra-rater kappa) below. The reliability estimate ranged from k = 0.19 to 0.91, with observed agreement ranging from 56.5 to 95.7%.

Adverse Events

At the completion of stage 2, 1 participant experienced an episode of syncope. She collapsed to the ground and immediately regained consciousness with no residual harm. Following this event, the researcher was mindful of asking participants to sit if they were feeling faint. No other adverse events occurred.

DISCUSSION

The primary aim of this study was to assess the intra-examiner and inter-examiner reliability of highly experienced chiropractors and recent graduates in their use of the Objective Synchronous Test (OST) as used in Advanced BioStructural Correction™ (ABC™) for testing

Table 3. Intra-rater Percentage Agreement

| INTRA-RATER PERCENTAGE AGREEMENT | RATER 1 | RATER 2 | RATER 3 | RATER 4 | AVERAGE |
|----------------------------------|---------|---------|---------|---------|---------|
| L5 | 95.7% | 73.9% | 60.9% | 56.6% | 71.8% |

Table 4. Intra-rater Kappa (95% Confidence Interval)

| INTRA-RATER KAPPA | RATER 1 | RATER 2 | RATER 3 | RATER 4 |
|-------------------|--------------------------|------------------------------|------------------------------|-------------------------------|
| L5 | 0.91 (0.74-1) P<0.001 | 0.49 (0.16 -0.82) P=0.012 | 0.20 (-0.19-0.61) P=0.349 | 0.13 (-0.28- 0.53) P=0.536 |

L5 for dysfunction. The intra-examiner reliability for the OST when testing L5 revealed a range of agreement from slight agreement for 1 inexperienced examiner and almost perfect agreement for 1 expert-level examiner.

For a test to be useful, it must perform consistently amongst different examiners and in different circumstances. A test must therefore have good reproducibility between different practitioners, or inter-examiner reliability. Tests that lack reliability are imprecise at best, and equivalent to guessing at worst. The results in this study, based on Fleiss' kappa, indicate that that examiners were reliable when using the OST for testing L5 with kappa values above 0.4 which has been identified as the minimum threshold for useful testing procedures in physical medicine. (21)

A secondary aim of this study was to compare the reliability estimates of chiropractors who were highly experienced ABC™ practitioners versus lesser experienced practitioners. As in other studies in the field, we found that there was no appreciable difference between these 2 groups when pooled. (18) These results may be initially surprising, given that experienced examiners are often presumed to have advanced skills gained over years of practice. In this study, we found no evidence that experienced chiropractors who practice ABC™ were more reliable overall than their less experienced counterparts. There was variation in the reliability of individual examiners, with 1 experienced examiner demonstrating almost perfect intra-examiner reliability. However, only 2 experienced ABC™ chiropractors were included in this study and the extent to which these results may be generalized is limited.

Strengths and Weaknesses of the Study

The study was modeled on the design used by Kmita. (14) Regarding study quality, every opportunity was taken to blind examiners from confounding factors that may have influenced their OST findings. Examiners were blinded to details of the patients' status as symptomatic or asymptomatic, and were blinded to all other clinical information. They were blinded to their prior findings for each participant and also the findings of other examiners during the study. Examiners were prevented from gaining information from study subjects regarding prior OST outcomes, and were also blinded to additional cues that may have influenced the results. The order in which examiners tested study participants was randomly varied at each stage of data collection. Tests were applied, interpreted and recorded

Table 5. Inter-rater Percentage Agreement

| INTER-RATER PERCENTAGE AGREEMENT | RATERS 1&2 | RATERS 3&4 |
|----------------------------------|------------|------------|
| L5 | 69.5% | 73.9% |

Table 6. Inter-rater Kappa Calculations (95% confidence interval)

| INTER-RATER KAPPA | RATERS 1&2 | RATERS 3&4 | RATERS 1-4 |
|-------------------|------------------------------|---------------------------|------------------------------|
| L5 | 0.42 (0.09 - 0.75) P=0.03 | 0.47 (0.12 - 0.83) P=0.02 | 0.49 (0.32- 0.66) P<0.001 |

according to current standards. The time interval between assessments was long enough and suitable for the variables being measured so that examiners raters would be unlikely to recall from memory alone. Categorical data were collected and analyzed with the kappa statistic and reported in conjunction with the standard error, observed agreement and expected agreement as suggested by Lantz. (22)

In relation to study applicability, 4 chiropractors who practice ABC™ with a range of experience were included as examiners.

This study included a sample of symptomatic participants representative of those seen in typical practice as recommended in the QAREL resource for the quality of diagnostic reliability studies. (23) Approximately 30% of the subjects had non-specific spine pain and the remainder were asymptomatic. This is representative of chiropractic practice where presentations for spinal pain and wellness (asymptomatic care) make up the majority of patient visits. (24) Prior studies have been criticized for using asymptomatic participants only and it is important to establish the reliability of these tests in a sample including symptomatic participants as this is more representative of typical practice. (24)

The sample size of 23 has been estimated by other authors to be adequate for this type of study. (14) It is possible that the theoretical assumption that the findings remain relatively stable may be flawed. In the latent period between phases, participants were allowed to relax in a reception area and use their phones. While this is a habit practice for many, it may be enough to cause bio-mechanical changes that would create confounding factors when attempting to measure intra-rater reliability. Similarly, merely by performing the testing procedure, particularly when done repeatedly as in this study, bio-mechanical changes may take place resulting in conflicting findings between examiners. The repeated (although gentle), shaking of the shoulders may have also been enough to cause bio-mechanical changes sufficient to confound the findings. However, were this is the case, we would not have seen such reliability between stage 1 & 2 as was demonstrated by examiner 1 and the clinical utility of the test would be questionable. In practice, other cues such as visual inspection, postural observation, sway, tissue palpation and patient response are available to a practitioner. Future research should evaluate other aspects of the ABC™ method evaluating reliability, validity and responsiveness. (25) Clinical practice is improved when

the strengths and limitations of testing procedures used by practitioners are well understood.

CONCLUSION

The OST was reliable in both intra- and inter-examiner reliability. Examiner 1 demonstrated almost perfect reliability between Stage 1 & 2. Other examiners had moderate to slightly positive kappa values on repeat testing, which shows reliability greater than chance alone. The kappa value for inter-examiner reliability across all examiners for L5 was 0.49, which is greater than the minimum threshold for clinical utility in physical medicine, arbitrarily set but generally agreed at 0.4. (21) Given that the OST is a primary part of the ABC™ method, greater emphasis on practitioner agreement is important in the training of ABC™ practitioners. The OST is of similar, or superior reliability at L5, to other commonly used diagnostic testing procedures used within chiropractic to determine the site of intervention.

REFERENCES

1. Triano JJ, Budgell B, Bagnulo A, Roffey B, Bergmann T, Cooperstein R, et al. Review of methods used by chiropractors to determine the site for applying manipulation. *Chiropr Man Ther* 2013;21(1):1. DOI: 10.1186/2045-709X-21-36
2. French SD, Green S, Forbes A. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. *J Manipulative Physiol Ther* 2000;23(4):231-8. DOI: 10.1067/mmt.2000.106101
3. Amarin-Woods K, Woods B, Moore C et al. Research priorities of the Australian chiropractic profession: a cross-sectional survey of academics and practitioners. *J Manipulative Physiol Ther* 2022;45:73-89.
4. Chiropractic Board of Australia, Annual Report 2021/22, [accessed online 1/12/22] <https://www.chiropracticboard.gov.au/About-the-Board/Annual-report.aspx>
5. Jutkowitz J. *Advanced BioStructural Correction, Instructional manual I*, 2009
6. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;88:1545-1602.

7. Sabnis AB, Chamoli U, Diwan AD. Is L5-S1 motion segment different from the rest? A radiographic kinematic assessment of 72 patients with chronic low back pain. *European Spine J* 2018;27:1127-1135.
8. Kirkaldy-Willis WH, Farfan H. Instability of the lumbar spine. *Clin Orthop Related Res* 1982;165:110-123.
9. Brismée JM, Atwood K, Fain M et al. Interrater reliability of palpation of three-dimensional segmental motion of the lumbar spine. *J Manual Manipulative Ther* 2005;13:215-220
10. Nolet PS, Yu H, Côté P et al. Reliability and validity of manual palpation for the assessment of patients with low back pain: a systematic and critical review. *Chiropr Man Ther* 2021;29: 1-20. DOI: 10.1186/s12998-021-00384-3
11. Haneline MT, Young M. A review of intraexaminer and interexaminer reliability of static spinal palpation: a literature synthesis. *J Manipulative Physiol Ther* 2009;32,:379-386. DOI: 10.1016/j.jmpt.2009.04.010
12. Merz O, Wolf U, Robert M, Gesing V, Rominger M. Validity of palpation techniques for the identification of the spinous process L5. *Man Ther* 2013;18:333-338. DOI: 10.1016/j.math.2012.12.003
13. Cooperstein R, Haneline MT. Spinous process palpation using the scapular tip as a landmark vs a radiographic criterion standard. *J Chiropr Med* 2007;6(3):87-93. DOI: 10.1016/j.jcme.2007.07.001
14. Kmita A, Lucas NP. Reliability of physical examination to assess asymmetry of anatomical landmarks indicative of pelvic somatic dysfunction in subjects with and without low back pain. *Int J Osteop Med* 2008;11(1):16-25. <https://doi.org/10.1016/j.ijosm.2008.01.003>
15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychosocial Measurement* 1960;20:37-46. 1960. <https://doi.org/10.1177/001316446002000104>
16. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*, Boston, MA; John Wiley & Sons; 2013.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;159-74. <https://doi.org/10.2307/2529310>
18. Seffinger MA, Najm WI, Mishra SI et al. Reliability of spinal palpation for diagnosis of back and neck pain. *Spine* 2004;29(19):E413 - E24. DOI: 10.1097/01.brs.0000141178.98157.8e
19. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Statistics Med* 1998;17:101-10 DOI: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e
20. Degenhardt BF, Snider KT, Snider EJ, Johnson JC. Interobserver reliability of osteopathic palpatory diagnostic tests of the lumbar spine: improvements from consensus training. *J Am Osteop Assoc* 2005;105(10):465-73.
21. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85(3), 257-268. <https://doi.org/10.1093/ptj/85.3.257>
22. Lantz CA, Nebenzahl E. Behavior and interpretation of the k statistic: Resolution of the two paradoxes. *J Clin Epidemiol* 1996;49(4):431-4. DOI: 10.1016/0895-4356(95)00571-4
23. Johnson C, Green BN. Public health, wellness, prevention, and health promotion: considering the role of chiropractic and determinants of health. *J Manipulative Physiol Ther* 2009;32(6):405-12. DOI: 10.1016/j.jmpt.2009.07.001
24. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010;63(8):854-61. DOI: 10.1016/j.jclinepi.2009.10.002
25. Khorsan R, Coulter I, Hawk C, Goertz C. Measures in chiropractic research: choosing patient-based outcome assessments. *J Manipulative Physiol Ther* 2008;5:355-375. DOI: 10.1016/j.jmpt.2008.04.007