

**TRADITIONAL CHINESE MEDICINE
TONGUE INSPECTION:
Development and evaluation of a
standardised inspection tool**

Minah Kim

BHlthSc (Acupuncture)/BA (International Studies),
MSc, MHlthSc (TCM)

A thesis submitted as partial requirement for the degree of
Doctor of Philosophy

Faculty of Science
University of Technology, Sydney

May 2009

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Abstract

Background: This author's previous MSc research concluded that traditional Chinese medicine (TCM) practitioners had poor levels of inter rater reliability (IRR) when observing tongue characteristics. This study suggested that IRR could be improved by developing operational definitions for the different tongue characteristics and a standardised protocol for observing the tongue. To this end this thesis documents the development and evaluation of a reliable tongue inspection tool and protocol.

Aim: To develop a reliable tongue inspection reference tool and protocol by examining different groups' tongue inspection interpretations.

Method: In addition to a reanalysis of the previous MSc study data (Study I), five studies were undertaken to obtain inter rater reliability (IRR) levels for different groups of subjects while developing and refining the inspection tool and protocol. For three within subjects experimental design studies (Study II, V and VI), IRR levels were measured before and after a formal training session of tongue inspection. For the other two studies (Study III and IV), a between groups comparison of IRR levels was undertaken both before (Study III) as well as after (Study IV) a session of training using the newly developed tongue inspection tool. All data collection sessions involved subjects viewing a variety of digital tongue images using a data projector or computer screen and the completion of a tongue inspection form. Study II measured the IRR levels of 28 Year 2 TCM students both before and after the standard TCM curriculum training of tongue inspection lecture. Study III compared the IRR levels of 22 Year 2 students who had not undertaken any formal training in tongue inspection to 23 Year 4 students who had previously had formal instruction in tongue inspection in their second year of TCM education. Study IV evaluated this same cohort of students some five weeks later following training with the improved inspection method using operational definitions, colour

charts and a standardised protocol. Study V, a within group repeated measures experimental design, measured IRR levels for both 21 inexperienced Year 2 TCM students while Study VI, used the same design, and involved 33 Year 4 TCM students. Both student groups had their IRR levels evaluated at baseline as well as after instruction in the newly developed tongue inspection tool. Statistical analysis involved assessing IRR with an agreement level of $\geq 80\%$ was used as a criterion for an acceptable level of reliability. In addition, t-tests and Pearson product correlation coefficients were also computed where applicable.

Results: Study I showed that using the new scoring method, the practitioners were poor at choosing the 'correct' answers and had a low level of agreement. Their performance did not improve or change across the two sessions.

For Study II, two characteristics achieved the criterion at both sessions (coat presence and body deviation) and there was one statistically significant increase in mean IRR value (body enlargement). Within both sessions 1 and 2, the initial and repeat viewing mean IRR values obtained for a given tongue example showed little change and in general were low and the mean percentage scores for the group of 20 subjects did not increase significantly from session 1 to session 2.

Study III demonstrated statistically significant differences for two tongue characteristics (coat presence and coat thickness) with coat presence achieving the criterion IRR for both groups. Both groups achieved overall IRRs ≥ 0.8 for the characteristics of coat presence, coat colour, coat peeling, body indent, crack presence and body deviation. For two of the 14 characteristics statistically significant IRR levels were achieved. In general, IRR profiles were quite similar.

Study IV revealed that the two year groups achieved the IRR criterion level for a similar number of characteristics (ten characteristics for Year 2 cf eight for Year 4) with coat texture, body colour, body enlargement and crack type failing to achieve the criterion for both groups.

No statistically significant differences were found between the two groups for any characteristics.

Study V demonstrated that the statistically significant changes were achieved for seven of the 14 characteristics from sessions 1 to 2 (coat colour; coat moisture; coat texture; body enlargement; body thorn; crack presence and crack type). This was reflected in the statistically significant increase in the group's mean score of 7%.

For Study VI, there were statistically significant increases for six of the 14 characteristics. These were coat peeling, coat texture, body enlargement, body indent, body thorns and body length. The mean percentage scores for this group increased significantly from session 1 to session 2 by 8%.

Conclusion: This series of studies on the reliability of tongue inspection has demonstrated that with appropriate instruction and the development of appropriate concrete operational definitions for the tongue characteristics and colour charts, significant improvements can be achieved in the reliability associated with identifying diagnostic characteristics. The use of this tongue inspection tool will advance education, clinical practice and research. TCM teaching institutions and individual practitioners should be encouraged to use and incorporate the tool into everyday practice. In addition further research can now assess whether the tongue diagnosis method is valid for specific TCM clinical diseases conditions.

Acknowledgements

There are many individuals who contributed towards the completion of the thesis. I would like to express my thanks to the Faculty of Science, Department of Medical and Molecular BioSciences and the College of Traditional Chinese Medicine for facilities and financial support; Professor Greg Skilbeck, Associate Dean (research) of Science faculty; Ms Rochelle Seneviratne, Science research and development officer; Professor Mark Tennant, Dean of University Graduate School and University Graduate School staff members; Student services Unit; Students' Association; and the Faculty of Science staff members and colleagues.

This work was supported by the Australian Postgraduate Award (APA) Scheme of Commonwealth Government.

I am very grateful for the 127 TCM undergraduate students who have participated in the study and completed all sessions involved in the data collection.

I am indebted to Dr Deirdre Cobbin, my principal supervisor, and Dr Chris Zaslowski, my co-supervisor, for their generous and invaluable supports and guidance throughout the candidature including allowing my access to their classes for data collection, assistance with statistical analysis and interpretation of the results, drafting and completion of this thesis.

I also would like to thank my family and friends for their endless love and supports. Thank you for being there for me.

Contents

Abstract	ii
Acknowledgements	v
Contents	vi
List of Figures	ix
List of Tables	xi
Chapter I: Introduction	1
Format of this thesis	3
Chapter II: Literature review	6
Previous studies	6
Need of operational definitions and review of the problems	9
Chapter III: Methods	17
Part I. Development of a standardised tongue inspection protocol	17
Identification of tongue characteristics	17
Definition of tongue topographic regions	18
Operational definitions	20
Development of the colour charts	26
Reproduction of the colour charts	32
Development of the tongue inspection form	32
Tongue inspection protocol	35
Part II. Study design, procedure and statistical analyses	38
Study II (2005)	38
Study III and Study IV (2006)	41
Study V and Study VI (2007)	44
Chapter IV: Results	48
Study I (MSc research study results)	48
Individual IRR	48
Overall IRR	51
Comparison of characteristics between the two sessions	51
Change in individual subjects' total scores between the two sessions	52
Correlation between the total scores for the two sessions	52
Summary of findings	53

Study II (2005)	55
Mean IRR for initial and repeat viewing of sessions	55
Mean IRR for initial and repeat viewing across two sessions	56
Mean IRR between the two sessions.....	57
Individual subject total scores in each session.....	59
Change in individual subjects' total scores between the two sessions.....	60
Correlation between the total scores for the two sessions	61
Group mean percentage scores.....	61
Summary of findings.....	62
Study III (2006 baseline measurement results).....	63
Individual IRR.....	63
Overall IRR	66
Comparison of the characteristics between the two year groups	67
Summary of findings.....	68
Study IV (2006 post training results)	69
Individual IRR.....	69
Overall IRR	72
Comparison of the characteristics between the two groups	73
Summary of findings.....	73
Study V (2007 Year 2 results)	75
Individual IRR.....	75
Overall IRR	78
Comparison of the characteristics between the two sessions.....	79
Individual subject total scores in each session.....	80
Change in individual subjects' total scores between two sessions	81
Correlation between subjects' total scores for sessions 1 and 2	83
Summary of findings.....	83
Study VI (2007 Year 4 results).....	85
Individual IRR.....	85
Overall IRR	88
Comparison of characteristics between the two sessions.....	89
Individual subject total scores in each session.....	90
Change in individual subjects' total scores between two sessions	92
Correlation between subjects' total scores for sessions 1 and 2	92
Summary of findings.....	93

Chapter V: Discussion	95
Tongue length	95
Difficulties of defining other characteristics.....	95
Tongue coat and body colour.....	96
Observation conditions	98
Study results.....	98
 Chapter VI: Conclusion and Recommendations	 110
 References	 113
 Appendix I: Tongue characteristics (Adopted from Kim 2003).....	 116
Appendix II: Subject consent form and information sheet	142
Appendix III: Instructions to down load Tongue examples slide and colour chart slide files.....	144
Appendix IV: Colour Charts 2006	147
Appendix V: Colour Charts 2007	149
Appendix VI: Order of presentation for tongue example for Study II, V and VI.....	151
Appendix VII: Study II Tongue examples.....	152
Appendix VIII: Study III and Study IV Tongue examples.....	153
Appendix IX: Study V and Study VI Tongue examples for sessions 1 and 2	163
Appendix X: Tongue inspection form 2005	173
Appendix XI: Tongue inspection form 2006	174
Appendix XII: Tongue inspection form 2007.....	178
Appendix XIII: Tongue inspection response answers for Study I to VI.....	182

List of Figures

Figure 2.1: A flowchart illustrating the process of document exclusion.....	6
Figure 2.2: Tongue topography based on Three Heaters.	15
Figure 2.3: Simple tongue topography based on body organs.	15
Figure 2.4: Complex tongue topography based on body organs - <i>Zang Fu</i>	15
Figure 3.1: Tongue characteristics and their various qualities commonly observed in tongue inspection.....	18
Figure 3.2: Tongue topography regions.....	19
Figure 3.3: A tongue example showing inspection regions of the tongue body and tongue coat.	20
Figure 3.4: Timeline of the data collection procedure.	44
Figure 3.5: Time line of the data collection procedure.....	46
Figure 4.1: IRR for each characteristic for individual tongue examples shown for session 1 (grey) and session 2 (black).....	50
Figure 4.2: Change in percentage scores for individual subjects from session 1 to session 2.	52
Figure 4.3: Correlation between total (out of 50) for the ten tongues (each out of 5).....	53
Figure 4.4: Mean IRR values from the initial and repeat viewing of the set of ten tongue examples at session 1(white: initial; pale grey: repeat) and at session 2 (dark grey: initial; black: repeat).....	57
Figure 4.5: Mean IRR values for individual tongue examples at session 1 (grey) and session 2 (black).....	58
Figure 4.6: Frequency distribution of total scores (%) achieved by the subject group at session 1.	59
Figure 4.7: Frequency distribution of total scores (%) achieved by the subject group at session 2.	59
Figure 4.8: Frequency distribution showing percentage change in total scores for the 20 observers from session 1 to session 2.....	60
Figure 4.9: Correlation between subjects' total scores for sessions 1 and 2.	61
Figure 4.10: Comparison of IRR values obtained for Year 2 (grey) and Year 4 (black) for the 14 characteristics shown for each of the eight tongue examples.	65
Figure 4.11: Comparison of IRR values obtained for Year 2 and Year 4 for the 14 characteristics shown for each of the eight tongue examples (grey: Year 2; black: Year 4; bold grey: repeated tongue Year 2; bold white: repeated tongue Year 4). .	71
Figure 4.12: Comparison of IRR values obtained at session 1 (grey) and session 2 (black) for the 14 characteristics shown for each of the eight tongue examples.	77
Figure 4.13: Frequency distribution of total scores (%) achieved by the subjects at session 1.	80

Figure 4.14: Frequency distribution of total scores (%) achieved by the subjects at session 2.	81
Figure 4.15: Individual subjects' change in percentage scored from sessions 1 to 2.....	82
Figure 4.16: Correlation between subjects' total scores for sessions 1 and 2.....	83
Figure 4.17: Comparison of IRR values obtained at session 1 (grey) and session 2 (black) for the 14 characteristics shown for each of the eight tongue examples.	87
Figure 4.18: Frequency distribution of total tongue scores (%) achieved by the subjects at sessions 1.	91
Figure 4.19: Frequency distribution of total tongue scores (%) achieved by the subjects at sessions 2.	91
Figure 4.20: Individual subjects' change in percentage scored from sessions 1 to 2.	92
Figure 4.21: Correlation between subjects' total scores for sessions 1 and 2.	93
Figure 5.1: Pale red 2006 body colour chart.	97
Figure 5.2: Red tongue 2007 body colour chart.....	97

List of Tables

Table 3.1: RGB codes for each coat colour categories used in colour chart I.....	28
Table 3.2: RGB codes for each body colour categories used in colour chart I.	29
Table 3.3: Munsell and RGB codes for each coat colour used in colour chart II.....	31
Table 3.4: Munsell and RGB codes for each body colour used in colour chart II.	31
Table 3.5: The structure and content of the tongue inspection form used for Study II.	33
Table 3.6: The structure and content of the tongue inspection form used for Studies III and IV. .	34
Table 3.7: The structure and content of the tongue inspection form used for Studies V and VI...	35
Table 4.1: Session 1 IRR values shown for the five main characteristics for each of the ten tongue examples.	49
Table 4.2: Session 2 IRR values shown for the five main characteristics for each of the ten tongue examples.	49
Table 4.3: t-tests comparing IRR for the two sessions for each characteristic.....	51
Table 4.4: Mean IRR values for the ten characteristics from the initial and repeat viewing of the set of ten tongue examples at session 1 and at session 2.	55
Table 4.5: t-tests comparing mean IRR values between the two sessions for each characteristic.	58
Table 4.6: IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall for the Year 2 subjects.	63
Table 4.7: IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall for the Year 4 subjects.	64
Table 4.8: t-tests comparing IRR between the two years for each characteristic.....	68
Table 4.9: IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall for the Year 2 subjects.	69
Table 4.10: IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall for the Year 4 subjects.	70
Table 4.11: t-tests comparing IRR between the two years for each characteristic.....	73
Table 4.12: Session 1 IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall.	75
Table 4.13: Session 2 IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall.	76
Table 4.14: t-tests comparing IRR between the two sessions for each characteristic.	79
Table 4.15: Session 1 IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall.	85
Table 4.16: Session 2 IRR values for the 14 characteristics for each of the eight tongue examples, shown individually and overall.	86
Table 4.17: t-tests comparing IRR between the two sessions for each characteristic.	89