UNDERSTANDING DIAGNOSTIC REASONING IN TCM PRACTICE: TONGUE DIAGNOSIS

Joyce K. Anastasi, PhD, DrNP, FAAN, LAc; Leanne M. Currie, RN, DNSc; Gee H. Kim, MA

Background • Traditional Chinese medicine (TCM) diagnosis is a complex multifaceted process that often yields multiple differential diagnoses and subdiagnoses.

Objectives • The aims of this study were to (1) understand cognitive strategies and diagnostic reasoning processes of TCM practitioners engaged in tongue diagnosis and (2) investigate TCM practitioners' diagnostic accuracy. Clinical decision making and problem solving frameworks served as a basis for this study.

Methods • Nine TCM practitioners verbalized their thinking processes via think-aloud protocols and were audiotaped while engaged in the cognitive task of tongue diagnosis. Protocol analysis was used to identify TCM diagnostic reasoning patterns. Diagnostic accuracy was assessed by participant self-report and via independent TCM expert judges.

Results • Protocol analysis revealed that TCM practitioners use systematic processes to arrive at diagnoses and that there were differences between novices and experts in both pre- and post-

adjustments after viewing case histories. Novices tended to use more descriptors and come to the diagnoses earlier. Experts tended to use higher-level intellectual processes when coming to their diagnoses and tended to use these terms earlier in the process. Correlations between practitioner self-assessment and judges' ratings of diagnostic accuracy were noted and corollary case history information improved diagnostic accuracies.

Conclusion • TCM practitioners use systematic reasoning patterns to determine diagnoses associated with evaluation of tongues. These processes are congruent with those observed in Western medicine whereby clinician reasoning involves a combination of analytical reasoning of domain knowledge and the use of exemplar patterns. An explicit understanding of TCM reasoning processes can inform clinical practice and education and will facilitate the development of supporting technologies and identification of best practices. (*Altern Ther Health Med*. 2009;15(3):18-28.)

Joyce K. Anastasi, PhD, DrNP, FAAN, LAc, is the Helen F. Pettit Professor and director of the Center for AIDS Research, **Leanne M. Currie, RN, DNSc,** is an assistant professor of nursing, and **Gee H. Kim, MA,** is a data manager, Center for AIDS Research, all at the Columbia University School of Nursing, New York, New York.

Traditional Chinese medicine (TCM) is a system of medicine with at least 2000 years of documented practice.[1,2] A growing body of research describes TCM practice,[3-5] treatment efficacy,[6-8] and cost-effectiveness,[3,4] but many aspects of this field warrant further research.[9-11] Despite the long practice history, little is known about the cognitive processes of practitioners while they are performing TCM diagnosis. The aim of this study was to understand tongue diagnosis, an aspect of TCM practice that involves evaluation and interpretation of physical characteristics of the tongue to derive a TCM diagnosis. This study explored 2 components of tongue diagnosis: diagnostic reasoning patterns and diagnostic accuracy. Understanding the process of tongue diagnosis is a first step in conceptualizing the underlying cognitive processes of TCM practitioners.

BACKGROUND

In the West, TCM practice has seen a recent increase in use, and a parallel increase has been seen in TCM research.[5,6] Researchers have examined the effect of TCM treatments on several chronic and acute symptoms including joint and muscle pain, headache, insomnia, anxiety, depression, menopause symptoms, digestive disturbances, and medication side effects.[12-14] A small number of studies have examined TCM practitioner demographics and practice characteristics including diagnostic processes and diagnostic accuracy.[15-21] These studies have found that TCM diagnostic activities are consistent across practitioners including biomedical or TCM interview,[10] pulse assessment, visual inspection, tongue diagnosis,[7,8] identification of multiple TCM diagnoses,[10,11] and use of landmark points for treatment of certain diagnoses.[22] Researchers also have found that diagnoses and treatment plans can vary between TCM practitioners[8,10,11,23] (See Box for descriptions of concepts and processes). To date, researchers have rarely focused on cognitive processes associated with TCM diagnoses; thus, the purpose of our study was to examine the TCM diagnostic process and to examine diagnostic accuracy using a TCM diagnosis as the reference standard.

Traditional Chinese Medicine Process

According to the normative, or prescribed, model of TCM

CONCEPTS AND PROCESSES

- Biomedical interview is based on evaluation of organ systems, whereas the TCM interview is based on the following characteristics: sleep, thirst, urinary output, bowel movements, appetite, digestion, energy level, sweating patterns, menses, and emotions. These features are evaluated as they relate to the patient's main complaint.[12]

- Pulse evaluation consists of palpating the radial pulse at 3 locations proximal to the wrist on each forearm. Each of these placements is related to a corresponding organ network. The pulse for each of these 6 locations is palpated by application of different degrees of pressure; light, moderate, and firm. Each pulse indicates the basic condition of the patient according to the 8-principle theory of TCM (yin and yang, excess and deficiency, heat and cold, and interior and exterior).[12]

- The patient assessment also includes an evaluation of the patient's entire demeanor, including brightness of the eyes, color of the nail beds, and general vigor.

- Qi, also written as chi and pronounced "chee," is considered a vital life energy that runs through meridians or "energy pathways" in the body and is a substance that exists in all living things; thus, the essence of qi can be taken into the body when one consumes plant and animal-based foods. The second substance, blood, is considered a "vital nutrient substance," which nourishes bodily tissues and organs and ensures adequate mental functions.[12] The third substance, body fluids, functions to moisten and nourish the body and includes saliva, gastric juices, joint fluids, tears, mucus, sweat, and urine. These 3 substances provide the foundation from which health and illness are understood.[9,12]

- The 5 characteristics are (1) element, including metal, earth, fire, water, and wood; (2) cycle of development, including germination, growth, transformation, reaping, and storing; (3) season, including fall, late summer, summer, winter, and spring; (4) climate, including wind, heat, dampness, dryness, and cold; and (5) emotion, including anger, joy, meditation, grief and melancholy, and fright and fear.[12]

- Tongue characteristics: body color refers to the substance of the tongue itself, that which lies beneath any coating. This color reflects the condition of the yin organs, the blood, and the nutritive qi (nutritive qi refers to the ability for qi to provide energy to a system). The shape of the body of the tongue is examined for width, length, cracks or ulcers, general texture, and movement. The tongue may be swollen or thin, short or long; the surface may be cracked or ulcerated; and the texture may be supple or stiff. Tongue coating is observed systematically from tip to the root of the tongue and reflects hot or cold aspects of tongue diagnosis. Tongue moisture provides an indication of the fluid quality in the person and will either be normal, overly moist, or overly dry. Movement is assessed to identify involuntary trembling, involuntary side-to-side motions, curling in any direction, or shifting to one side. Vitality, also known as tongue spirit, denotes the apparent liveliness of the individual and is an indicator of the presence of an underlying pathological condition. TCM practitioners describe vitality by characterizing the "Tongue of Life," which has a very vital color, and the "Tongue of Death," which has a "withered and dark" appearance. Both of these characteristics of "vitality" are most important when visualizing the root of the tongue, the region where the tongue emanates from the throat.

practice, patient evaluation consists of 4 phases: (1) interview; (2) physical assessment including palpation of radial pulse, visual evaluation of the tongue, and general physical exam; (3) generation of a TCM diagnosis; and (4) formulation of a treatment plan (Box).[12,24] At the conclusion of the interview and assessment, the practitioner generates 1 or more TCM diagnoses. Once the TCM diagnosis is established, the treatment plan is determined and can include acupuncture, moxibustion, massage, herbs, and dietary recommendations.

TCM theory posits that health is determined by a balanced flow of 3 substances: qi, blood, and body fluids[12] (Box). Illness is identified via the 8-principle theory with the following 4 paired components: (1) hot and cold, (2) excess and deficiency, (3) interior and exterior, and (4) yin and yang. Using these 8 principles, the TCM practitioner identifies imbalances related to the patient's qi, blood, or body fluids. TCM diagnosis is also based on 5 organ networks: the lungs/colon, spleen/stomach, heart/small intestine, kidneys/bladder, and liver/gallbladder. These networks are then associated with 5 characteristics: element, cycle of development, season, climate, and emotion.[12] Each organ network is capable of incurring TCM pathogenic factors arising from imbalances in the 1 or many of the 5 characteristics. Pathogenic factors include wind, heat, dampness, phlegm, dryness, and cold. When organ networks are in a state of imbalance, such as excess or deficiency in qi, blood, or fluids, the patient may demonstrate symptoms, disease, injury, or stress.[12]

Tongue Evaluation

Tongue evaluation dates from the Shang Dynasty (1600 BC to 1000 BC) and consists of visually inspecting the tongue body for vitality, color, shape, moisture, and movement and assessing the tongue coating for color, thickness, distribution, and characteristics at the root.[13-15] According to TCM theory, the tongue provides a geographic map of organ systems; characteristics of the tongue in each of these areas provide information critical to the TCM diagnosis (see Box). Figure 1 shows corresponding tongue locations for organ systems,
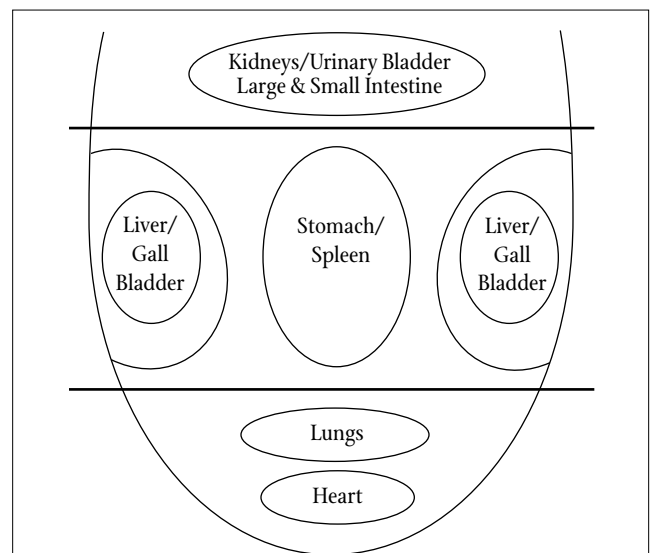

FIGURE 1 Corresponding Locations of TCM Features on the Tongue

and Table 1 shows assessment components for TCM tongue evaluation with select corresponding TCM patterns.

CLINICAL DECISION MAKING AND PROBLEM SOLVING

As with Western medicine, TCM practitioners are faced with clinical problems. Clinical decision making in Western medicine has been described as a delicate process of organizing and interpreting domain information (ie, information that clinicians use to understand and guide their practice) and patient information. Such clinical decisions are often made in situations of uncertainty or under time constraints.[16] Several models of professional problem solving have been put forth, including the hypothetico-deductive model, a normative model, popularized by Karl Popper, in which the scientific process is characterized as theory or hypothesis proving.[17-19] The hypothetico-deductive model posits that scientists use a systematic process to establish a theory and then set about falsifying that theory.[18] Extended to the clinical setting, the hypothetico-deductive model asserts that clinicians use available evidence to logically identify a diagnostic hypothesis which is then systematically confirmed or refuted.[17]

Observations of clinicians in their natural environments, however, have demonstrated that the hypothetico-deductive model does not fully characterize the clinical decision-making process.[20] Rather, since health-related problems are complex and may have multiple causes, it is unlikely that a single diagnostic hypothesis will fit each patient picture.[21] Researchers in this field generally concur that 2 processes are involved in medical decision making: exemplar-based reasoning and analytic reasoning. Exemplar-based reasoning is used when clinicians use exemplars or similar cases that are well-represented in the problem domain (ie, pattern matching).[20,21,25] Exemplar-based reasoning is different from the process described in the hypothetico-deductive model because exemplar-based reasoning does not possess an overt and logical process. Analytic reasoning is the systematic process of categorization whereby clinicians search for and identify relevant features and then systematically weigh the relevance of each feature against their interpretation of domain information (ie, the individual's representation of the knowledge in their domain).[20,25] Both of these processes can be used during diagnosis of an individual case.

Expertise also plays a role in the accuracy of decision making and problem solving, although the exact influence is unclear.[20,21,25] The function of expert performance is well characterized in Western healthcare with the goal of promoting safe and effective practice.[26-32] In general, expertise is acquired after 5 to 10 years of continuous practice for a specific task (at least 4 hours per day), and the development of expertise depends on exposure to enough cases, including a wide variety of rare and unusual cases.[29,33] Experts tend to match patterns (exemplars) from previous experiences, effectively using shortcuts (or *heuristics*), which facilitate more rapid decision making in situations of uncertainty.[21,29,34] Although most shortcuts are accurate, some may result in erroneous judgment.[17]

Use of Visual Information for Decision Making

Research in healthcare domains that use visual information, such as radiology, pathology, dermatology, and wound care, has focused on understanding the cognitive processes associated

TABLE 1 Aspects of Tongue Assessment and Corresponding TCM Factors				
Aspect	Health	Illness	Clinical Significance	8 Principles
Tongue Characteristics				
Vitality/tongue spirit	"Tongue of Life": vibrant and vital	"Tongue of Death": withered and dark	Denotes state of liveliness irrespective of any coincident pathological signs	
Body color	Pink	Red, pale, bluish	Denotes condition of yin organs, blood, and nutritive qi	Yin/yang Cold/hot
Body shape	Supple, neither swollen nor thin	Thin, swollen, long, short	Denotes condition of yin organs, blood, and nutritive qi	Deficiency/excess
Moisture	Slightly moist	Dry, overly moist	Denotes condition of body fluids	Cold/hot Yin/yang
Movement	Centered, even, symmetrical, calm	Quivering/moving	Denotes clinical phenomenon, eg, quivering could indicate a deficient condition of blood such as Bell's palsy	Deficiency/excess Wind
Coating Characteristics				
Color	Clear-white	White, yellow, grey	Hot/cold conditions	Cold/hot
Thickness	Thin	Very thin, thick, slippery	Hot/cold conditions Interior/exterior conditions	Deficiency/excess
Distribution	Uniform	Uneven	Interior/exterior conditions	Interior/exterior
Root	Thin	Coated	Exterior pathogen	Deficiency/excess

with visual pattern recognition and mental representation of the domain-related images.[35-39] For any domain, the methods by which visual information is interpreted must be understood in order to anticipate potential biases, errors, and visual illusions that might yield ineffective or erroneous decisions.[35,40-42] Researchers have noted that dermatology experts were more accurate when they made rapid diagnoses[43] but that misleading case scenarios could make even very obvious lymph node enlargement invisible to physicians.[44] Other researchers have found that provision of sequential photographs improved correlation of prediction of wound infection between novices and experts.[36] Research with pathologists identified that, in general, expert pathologists reported more findings, verbalized more cause-and-effect relationships, and demonstrated longer reasoning chains than novices.[45,46] In summary, a large body of research has explored decision making related to use of medical information. Research related to decision making in visually intensive domains may identify best practices and the best way to manage visual and corollary information.

METHODS

This was a descriptive study that used think-aloud protocols and protocol analysis to examine the cognitive processes and diagnostic accuracy of TCM practitioners as they performed tongue diagnosis.

Participants and Recruitment

Research participants were TCM practitioners of varying levels of experience. The participants were identified from local acupuncture clinics via flyers and by word of mouth. In addition, practitioners were purposefully sampled to capture 2 levels of experience: novice and expert. Novices were those with less than 5 years' experience (n=5), and experts were those with more than 5 years' experience (n=4). Inclusion criteria included National Certification Commission for Acupuncture and Oriental Medicine credentialing with the New York State licensing board and agreement to participate. IRB approval was obtained prior to the initiation of the study. All participants were provided with an overview of the research study protocol and then were asked to read and sign a consent form before the session began.

Study Setting

The research was conducted in a laboratory environment. This was a quiet location where no interruptions would occur and in which the researchers would have direct control over the participants' experience. All interviews were carried out at Columbia University School of Nursing (CUSN) Center for AIDS research (CFAR), an acupuncture suite and research center located within CUSN.

Procedures

After informed consent was obtained, a demographic survey was rendered, and the participants were instructed about the study protocol. Each session was scripted to ensure consistency across sessions. Four high-quality photographs of patients' tongues and associated case histories were obtained from a standard TCM education textbook by Kirschbaum.[13] The high-quality photographic prints were cut from the textbook and then laminated in clear plastic to ensure that they were protected and thus consistent for each participant's evaluation. The researchers and one additional TCM expert reviewed the case histories for content validity.

We randomly ordered the presentation of the tongue images to the participants based on the sequence of participant entrance into the study, and we used counterbalancing, a process of reordering the tongue photographs between participants, to control for the potential confounding influence of previous exposures to materials.[47] For example, participant 1 received the scenarios in the order of A-B-C-D, whereas participant 2 received the scenarios in B-C-D-A order.

Think-aloud protocols were used to identify TCM diagnostic reasoning patterns. Participants were asked to think aloud or verbalize their thought processes as they made their TCM diagnosis.[48-50] We also asked participants to use a felt-tipped pen to outline areas of interest on a piece of plastic that was placed over the image. The act of drawing was used to ensure that there was no time for introspection and potential amendment of actual thinking processes.[33] In addition, we provided a training session with a tongue photograph not used in the study to ensure that the practitioners would become comfortable with their task before the session commenced.[33]

Participants were asked to review the tongue photograph, after which they were asked to state their diagnosis(es) and indicate their level of confidence for that diagnosis using a 0-10 visual analog scale. After the level of confidence was declared, participants were provided with a case history including a description of the patient's pulse and constitutional findings. After reading the case history, participants were invited to adjust their initial diagnosis(es) if desired and to indicate their level of confidence of the revised diagnosis using the 0-10 visual analog scale. Each session was audiotaped, transcribed by a professional transcription service, reviewed by the researchers, and imported into NVivo (QSR International Party Ltd, Doncaster, Victoria, Australia), a computer application used to code descriptive and qualitative data.[51] On completion of the session, which lasted approximately 1 hour, each individual was compensated with $100 USD for his or her time.
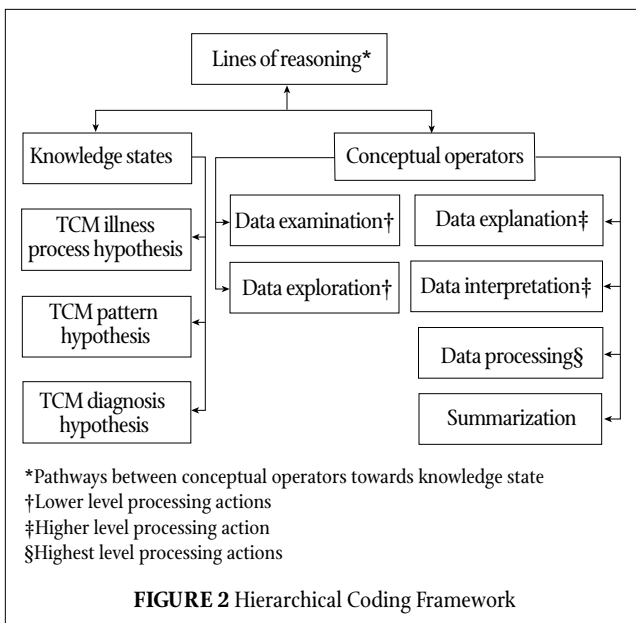
Coding Framework

Because our goal was to understand diagnostic processes in the health domain, we used a framework that was developed by Hassebrock and Prietula for characterizing diagnostic processes of physicians.[33] This framework uses 3 components to describe diagnostic reasoning: knowledge states, conceptual operators, and lines of reasoning.

A knowledge state is a characterization of domain understanding at a particular time point: for example, a disease hypothesis, a causative (pathophysiological) hypothesis, or a category (grouping) hypothesis. Each knowledge state can be modified by

conceptual operators, which are "building blocks" used by humans during reasoning and inferencing.[33,52] Conceptual operators include actions such as data examination, comparison of data to a normal presentation, and actions for diagnostic hypothesis generation. These actions indicate high- or low-level reasoning processes such as "read," a low-level process, or "evaluate own ability," a high-level process. Lines of reasoning represent the problem solvers' mental model of the "interrelationships associated with a particular problem." The framework was used to characterize the information that the clinicians know (knowledge state), the actions that they use to come to this knowledge (conceptual operators), and the relationships between knowledge components toward problem solving (lines of reasoning).

Crowley et al modified the Hassebrock and Prietula coding framework to capture the diagnostic processes of pathologists.[45,46] Crowley's modifications included regrouping conceptual operators considered redundant or similar. For example, Crowley added the conceptual operator artifact to capture any event in which the pathologist recognized a flaw in the visual information and in which the information would likely be excluded from the evaluation. Both coding frameworks were examined for suitability to characterize TCM practitioner behavior. For the final framework, we used the Hassebrock and Prietula framework as the foundation with the following adjustments: (1) the conceptual operator code, "identify artifact," from the Crowley framework was used with Dr Crowley's permission; (2) specific TCM-based characteristics were added to the conceptual operator code, "data examination"; (3) "data explanation" and "data exploration" were modified to reflect TCM-based activities; (4) "hypothesis generation" and "hypothesis evaluation" were subsumed under the conceptual operator code "data interpretation"; and (5) "discrepancy processing" and "meta-reasoning" were subsumed under the category "control processing." Figure 2 shows the hierarchical coding framework used for this study, and Table 2 shows the coding framework.

*Pathways between conceptual operators towards knowledge state
†Lower level processing actions
‡Higher level processing action
§Highest level processing actions

FIGURE 2 Hierarchical Coding Framework

Data Analysis

Protocol analysis was used to code the think-aloud verbal data that were obtained during the think-aloud sessions. Protocol analysis is a rigorous methodology used for eliciting verbal reports of thought sequences and is seen as a valid source of data on thinking.[35,40,45,52] The verbal data were coded using NVivo software, first by searching for key words and then by reviewing the transcripts multiple times to identify themes and general participant processes. The research team reviewed coding at regular meetings to discuss incongruities or issues. Any issues were solved by consensus. Once the coding was completed, 2 reports were generated from NVivo: one to tally the frequency of conceptual operator use by each study participant and one that provided the sequence of conceptual operator use for each study participant. Each file was then exported to MS Excel. In addition, the stated diagnoses that were asked for pre– and post–case history were obtained from the transcripts for each participant. Frequency of conceptual operator use was performed for each participant with a subanalysis by novices and experts. *T*-tests were performed to examine differences between conceptual operator use for novices and experts. Content analysis was performed to identify the sequence of conceptual operators, which permitted identification of lines of reasoning, including notable thinking patterns and processes associated with TCM tongue diagnosis.

Assessment of Diagnostic Accuracy

To assess for diagnostic accuracy, study participants' pre– and post–case history diagnoses were compared to the textbook diagnosis associated with the tongue photograph that was validated by TCM experts. We validated the textbook diagnosis by the following process: (1) 2 TCM judges were provided with the tongue photographs, the associated case histories, the descriptions of the associated pulses, and the corresponding textbook diagnoses, and then (2) each judge determined the degree to which the textbook diagnosis was accurate and then they clarified, added to, or adjusted the diagnosis accordingly. Since similar TCM diagnoses can be expressed in different ways—eg, deficient heat is equivalent to yin deficiency and interior wind is equivalent to liver wind—any discrepancies between the TCM judges' diagnoses and the textbook diagnoses were discussed until both judges came to agreement on a gold-standard TCM diagnosis or diagnoses for each tongue photograph. Based on the validated diagnosis, both TCM judges rated the accuracy of the TCM practitioners' diagnoses for each of the tongues on a scale of 0 to 10, with 10 being the most accurate. To minimize bias in rating participants' TCM diagnostic performance, the cases were presented to the TCM judges in a random order. Both judges' scores were assessed, and where there was a discrepancy, the judges conferred until consensus was reached. Percentage change between pre– and post–case history and correlations were calculated.

RESULTS
Demographics
Nine TCM practitioners (8 females) with a mean age of

TABLE 2 Coding Framework: Knowledge States Definitions and TCM Conceptual Operators

Knowledge States	Hassebrock and Prietula Definitions	TCM Tongue Diagnosis Interpretations
Disease hypotheses	Physical defects that constitute standard instances of disease	TCM patterns that indicate a TCM diagnosis. Not all characteristics of TCM diagnosis would necessarily constitute a physical defect
Pathophysiological hypotheses	Abnormal states or conditions that result from disease	TCM rationale for physical cue. Again, not all TCM processes have an underlying "pathophysiology"
Category hypotheses	Super-ordinate categories of either disease or pathophysiological hypotheses–"organization of various patterns of data, disease and pathophysiological anomalies on the basis of shared features and properties"	Super-ordinate categories of either TCM patterns or TCM rationale–"organization of various patterns of data, illness models, and deviations from normal from a TCM perspective on the basis of shared features and properties"

TCM Conceptual Operators Categories and Category Definition

Subcategories	Actions
1. Data Examination: Identification and examination of visual cues and other information. Uses low-level reasoning	
History	Read†; identify historical cue*
Identify visual cue*	Identify tongue body—color, shape, quality; relate diagnosis to history*; identify tongue spirit, vitality, coating—color, root, distribution, quality
Identify normal structure	Identify normal structure†
Compare	Compare to normal*; compare to expected*; compare to other cases*; compare findings in other areas*; determine degree of abnormality*
Identify artifact	Identify artifact* (an artifact is a flaw in the visual presentation)
Identify location	Anatomic location†; TCM location
2. Data Exploration: Higher order interpretation leading to abstraction, such as the assignment of severity and certainty	
	Assign-significance*; determine severity*; evaluate certainty of findings*; note absent data*; summarize findings*
3. Data Explanation: Explanation of visual cues denoting domain knowledge and understanding of theoretical model of domain	
	Infer TCM pattern; signs and symptoms; infer TCM substance; infer 8 principles condition
4. Data Interpretation: Higher level reasoning indicating synthesis of visual cues to support domain knowledge	
Hypothesis evaluation	Confirm with present finding; refute with present finding*; confirm with absent finding*; hypothesis certainty by visual analog; reject hypothesis*
Hypothesize associated data	Hypothesize associated data†; explain present finding with hypothesis*; search for features to discriminate*; evaluate certainty of hypothesis*
Generate Hypothesis	State hypothesis*
5. Control Processes: Control processes include meta-reasoning, discrepancy processing and other higher-level control processes. Meta-reasoning includes actions involved in the evaluation of the progress or quality of the reasoning process itself	
	Discrepancy recognition†; discrepancy revision†; discrepancy rationalization†; ignore discrepancy†; meta-reasoning—evaluate own ability; meta-reasoning—assess difficulty; meta-reasoning—state plans*
6. Summarization	Repeat significant facts†; repeat hypothesis†

*Items from Crowley Framework; †Items from Hassebrock & Preitula.

45.4±12.4 years participated in the study. The practitioners' TCM experience ranged from 6 months to 25 years with a mean of 6.7 years' experience, and 7 of the 9 practitioners had a master's degree or higher.

Conceptual Operators

During the think-aloud protocols, a total of 7715 conceptual operators were generated. Each practitioner used between 115 and 318 conceptual operators for each tongue. There was no dif-ference between overall mean conceptual operator use between novices and experts (208±44.23 for novices and 222±45.7 for experts; t=0.92, P=.364). We also examined the number of conceptual operators used for each tongue photograph and saw little variation in conceptual operator use except for a lower number of conceptual operators used by novices for tongue A.

When conceptual operator use was examined at the individual conceptual operator level, there were differences in use between novices and experts for several conceptual operators (Table 3). In

TABLE 3 Frequency of Select Conceptual Operators Used for Novices and Experts During the Diagnostic Process

Conceptual Operators	Expert Mean ± SD	Novice Mean ± SD	*P*
Data Examination			
Compare to normal	1.38 ± 1.59	1.50 ± 1.70	.82
Compare to expected	0.38 ± 0.72	0.80 ± 1.15	.21
Compare to other cases	1.63 ± 1.15	1.75 ± 2.17	.84
Compare to findings in other areas	1.38 ± 1.50	0.60 ± 1.05	.08
Determine degree of abnormality	0.56 ± 0.89	0.65 ± 1.23	.81
Identify artifact	0.50 ± 1.51	1.45 ± 2.24	.16
Data Exploration			
Assign significance	1.69 ± 1.82	1.05 ± 1.93	.32
Determine severity	0.44 ± 0.73	0.90 ± 1.97	.38
Evaluate certainty of finding	1.69 ± 1.74	0.80 ± 1.11	.07
Note absent data	1.63 ± 2.09	2.30 ± 2.45	.39
Data Interpretation			
Explain present finding with hypothesis	0.38 ± 0.62	0.10 ± 0.31	.09
Search for feature to discriminate	0.75 ± 1.13	1.15 ± 0.67	.19
Evaluate certainty of hypothesis	1.63 ± 1.20	0.85 ± 1.14	.06
Control Processes			
Discrepancy recognition	4.63 ± 3.34	3.75 ± 2.81	.40
Discrepancy revision	3.63 ± 2.68	1.55 ± 1.70	.01*
Discrepancy rationalization	3.63 ± 2.25	3.05 ± 2.76	.51
Ignore discrepancy	0.13 ± 0.34	0.30 ± 0.57	.29
Meta-reasoning–evaluate own ability	4.00 ± 3.08	3.85 ± 5.64	.92
Meta-reasoning–assess difficulty	1.38 ± 1.67	0.90 ± 2.02	.46
Meta-reasoning–state plans	1.06 ± 1.61	0.55 ± 1.00	.25
Summarization			
Repeat significant facts	0.94 ± 0.93	0.95 ± 1.28	.97

*Significant at *P*<.05.

general, novices used more low-level conceptual operators to describe the concrete features of the tongue photos, whereas the experts used more higher-level conceptual operators to describe interpretive processes. Of the data examination conceptual operators used, novices used more terms in all categories except when they compared the tongue image to the code "findings in other areas." Novices also tended to identify more artifacts than experts when examining the tongue photographs. When the participants were exploring the tongue photographs and corollary information, experts assigned significance and evaluated the certainty of their findings more commonly than novices, but novices evaluated salience (congruence of overall diagnostic conclusions) and noted absent data more frequently. In the higher-level functions, experts evaluated the certainty of their findings and explained the findings more commonly than novices, whereas novices searched for features to discriminate more commonly than experts did. A similar pattern was observed in the use of control processes: experts were more likely to use higher-level control processes than novices, except for the control process "ignore discrepancy," which novices used more commonly.

Diagnostic Reasoning Patterns—Lines of Reasoning

To identify lines of reasoning, the data were analyzed in 2 sections: pre–case history reasoning (first diagnostic hypothesis) and post–case history reasoning (final diagnostic hypothesis).

*Pre–Case History Diagnostic Reasoning*

As described above, during the pre–case history diagnostic reasoning phase, the practitioner was asked to describe the tongue and interpret the visual information in the manner in which he would normally conduct a tongue assessment and then state a first diagnostic hypothesis. All but one practitioner began his or her diagnostic processes with a description of the tongue body for all 4 of the tongue photographs examined; the ninth practitioner, an expert, started with a description of the tongue coating for 1 tongue only. For 23 of the 36 tongue assessments (64%), the practitioners described the tongue bodies via this sequence: body color, body shape, body quality. Practitioners used 2 of the 3 operators (color, shape, quality) in varying pairings for the remaining descriptions. Descriptions of tongue coatings were less consistent: practitioners described 9 of the 36 tongues (25%) via the sequence coating color, coating at root, coating distribution, and coating quality; 20% described the pair coating color and coating quality; 11% described the pair coating distribution and coating quality. For the remaining tongue descriptions, practitioners used various sequences, including coating at root, coating distribution, coating quality (8%) and coating color, coating distribution, and coating quality (8%). The pairing coating distribution and coating quality occurred in 53% of the description sequences.

The descriptions before the first diagnostic hypothesis also contained several pairings of conceptual operators. Practitioners frequently paired TCM pattern with specific patient characteristics. For example, a TCM pattern of spleen qi deficiency was paired with a patient characteristic such as "got a lot of emotional issues." A second notable pairing pattern was between TCM substances and TCM 8 principle conditions (eg, qi plus deficiency). None of these patterns were associated with the sequences of viewing the tongue photographs.

For novices, descriptions of the tongue body and coating were most commonly followed by the following sequence: (1) description of anatomical location (eg, on the sides of the tongue); (2) description of TCM location (eg, liver, heart, gall bladder area); (3) description of TCM patterns (eg, liver qi stagnation, liver fire, damp heat in the liver); (4) description of associated characteristics of the individual (eg, poor circulation); (5) TCM substance (eg, blood, qi); and (6) TCM 8 principles condition (eg, deficiency) (Figure 3).

Experts used comparison descriptors earlier in the descriptive sequence than novices. Novice 1 did not use comparison operators for 2 of the tongues; Novice 2 used comparison conceptual operators consistently later than experts. Experts 3 and 4 with more than 10 years of experience used comparison operators consistently much earlier than those practitioners with less experience. The number of comparison conceptual operators was widely varied between the practitioners (Table 3).

Prior to declaring the first diagnostic hypothesis, the

FIGURE 3 Sequence of Conceptual Operators Used by Novices and Experts Prior to First Diagnosis
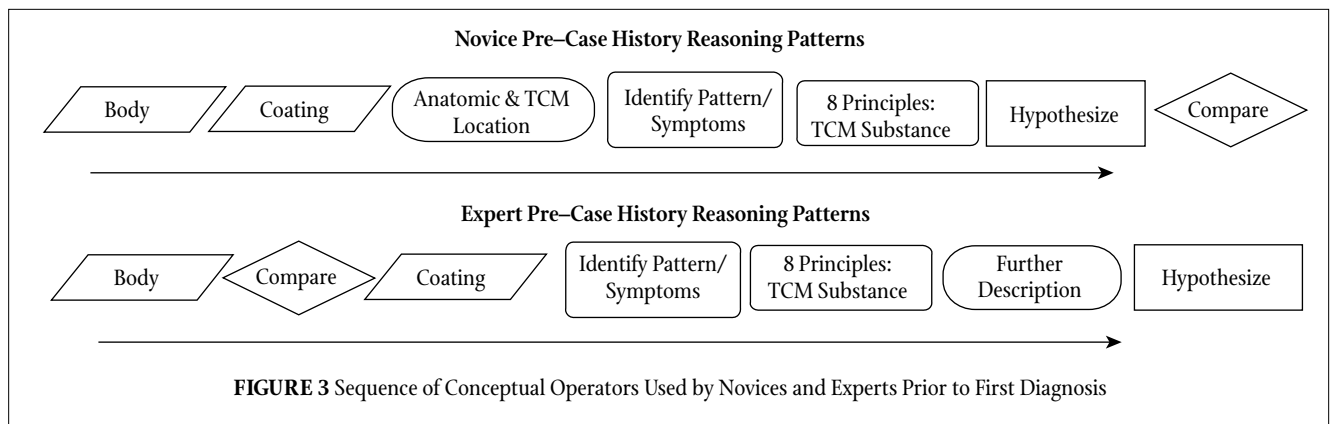
experts' descriptions of the tongue body were most commonly followed by (1) use of comparison conceptual operator, (2) description of coating, (3) description of TCM patterns, (4) description of associated symptoms, (5) TCM substance, and (6) TCM 8 principles condition. Figure 3 shows the general sequence of conceptual operators noted in the pre–case history (first diagnostic hypothesis) phase of the analysis for novices and experts.

Post–Case History (Final Diagnostic Hypothesis) Diagnostic Reasoning

After the case history was provided to the practitioners, they were asked to continue describing their interpretations of the tongue toward finalizing their diagnoses. Both novices and experts explained the characteristics of the tongue and explored the tongue photograph again. Novices tended to explain and then explore, whereas experts tended to explore and then explain. Both groups would recognize discrepancies in earlier interpretations (prior to receiving case history information), but novices were more likely to rationalize the discrepancy without making any adjustments to their initial diagnoses. In contrast, experts would revise their diagnoses once a discrepancy was recognized. Experts used conceptual operators related to higher-level meta-reasoning processes earlier in the reasoning process. These conceptual operators included evaluation of the practitioner's own ability, the assessment of difficulty, and the statement of planning for care. The general sequence observed for novices and

experts between receiving case history information and coming to a final diagnostic hypothesis is depicted in Figure 4.

Diagnostic Accuracy

Each practitioner generated 2 distinct certainty ratings for each diagnosis, one before and one after reviewing the case scenario; therefore, 72 ratings were made. Each practitioner generated between 2 and 5 TCM diagnoses per tongue case. All practitioners successfully identified at least a portion of the gold-standard diagnosis. For all but 2 of the practitioners, post-scenario self-assessment ratings were higher than pre-scenario assessment. Table 4 summarizes the mean certainty ratings by participants (novices and experts) and by TCM judges and correlations between participants and judges. Although the small sample size limits the interpretation of the mean scores, the data are presented here to display the trends that were observed in our study. Overall, practitioners made relatively accurate diagnoses as evaluated by the judges (6.58±1.89 overall on a 0-10 scale) on average. In addition, there was improvement across groups when additional corollary information was provided (7.22±1.48). Interestingly, for novices, the percentage increase (percent change between pre- and post-corollary information) in the self-rated accuracy score was larger than the increase in the judge-rating, perhaps indicating increased self-confidence in their diagnostic ability as compared to the gold standard.

FIGURE 4 Post–Case History Diagnostic Reasoning Patterns
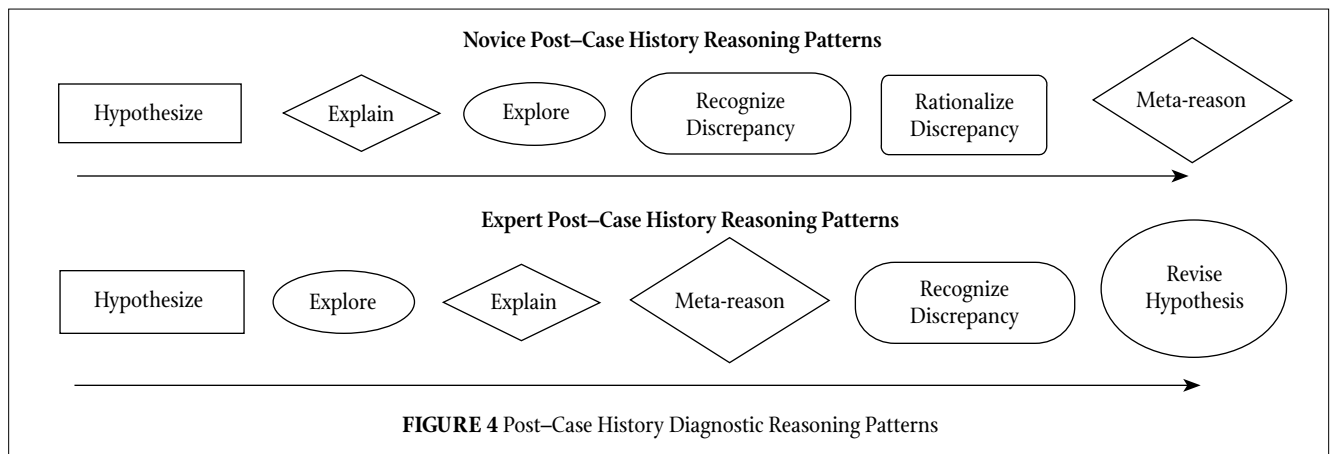
Correlations between participants' self-ratings (whether novice or expert) and judges' ratings were not statistically significant. In general, however, the correlation between the judges' ratings and experts' self-ratings was stronger than the correlation between the judges' ratings and the novices' self-ratings (Table 4).

TABLE 4 Correlation Between Self-rated and Judge-rated Accuracy for Tongue Diagnoses			
	Pre–case History	Post–case History	% Change
Novice (n = 5)			
Self-rated accuracy	7.23 ± 1.39	8.45 ± 1.65	16.8%
Judge-rated accuracy	6.83 ± 1.65	7.15 ± 1.23	4.7%
Correlation between judge and self-rating	r=0.23, P=.34	r=0.17, P=.48	
Expert (n = 4)			
Self-rated accuracy	7.44 ± 1.49	8.09 ± 1.93	8.7%
Judge-rated accuracy	6.28 ± 2.28	7.31 ± 1.78	16.4%
Correlation between judge and self-rating	r=0.26, P=.33	r=0.35, P=.19	

DISCUSSION

This descriptive study used think-aloud protocols and protocol analysis to identify the reasoning patterns of TCM practitioners as they evaluated photographs of tongues and arrived at TCM diagnoses. We also examined the diagnostic accuracy of varying levels of TCM practitioners as compared to a gold-standard diagnosis for 4 tongue photographs. To our knowledge, this is the first study to report on the reasoning patterns of TCM practitioners while they were making tongue diagnoses. Our results have revealed that, in general, tongue diagnosis is a multifaceted process during which TCM practitioners systematically evaluate the tongue and use consistent processes to determine diagnoses. Experts use more high-level reasoning actions, and novices use more low-level reasoning actions. Because of the structured task associated with this research, the analysis was divided into 2 segments, reasoning processes pre- and post-review of case history. We have identified distinct reasoning patterns before and after reviewing case history information, and we have identified differences between novices and experts in both pre–and post–case history thinking processes.

A coding framework was created by adapting frameworks developed by Hassebrock and Prietula to evaluate medical diagnostic processes and by Crowley et al for the visually intensive field of pathology diagnoses. Our framework was useful for capturing concepts related to both domain knowledge and the clinical reasoning process. In addition, protocol analysis of the coded verbalizations provided an efficient method to describe the diagnostic reasoning process of clinicians. Participants were asked to use a felt-tipped pen to outline areas of interest on a piece of plastic that was placed over the image. This ensured that there was no time to ponder and to potentially modify the thinking processes. This is consistent with Hassebrock and Prietula's conten-

tion related to think-aloud protocols in which a focused task can provide unhindered access to the thinking processes of the individual. The use of NVivo software facilitated the analysis of this large volume of descriptive data.

Different characteristics were noted between novice and expert TCM practitioners in both the process of arriving at an initial diagnosis and the process of confirming or refuting the initial diagnostic hypothesis. Pre–case history diagnostic reasoning patterns showed that both novices and experts used a systematic series of descriptors to assess the tongue body followed by the tongue coating. This is an interesting finding because although the sequence that we observed might be a function of rote training, which would strongly influence a novice's behavior, it demonstrates that this process is perpetuated in experts. That is, the normative model may actually be used in applied practice, which is inconsistent with current research related to medical reasoning in which the normative model is the hypothethico-deductive model.

Novices, in general, used more conceptual operators than the experts, which may be a function of the development of exemplar patterns that will be used as they become more experienced. Novices were more likely to use low-level conceptual operators to describe the anatomic location in association with the TCM location, which also might be related to the process of assimilating domain knowledge as is seen in less experienced practitioners.

During the pre–case history phase, novices more commonly followed a systematic sequence that involved use of low-level operators to describe visual characteristics of the tongue. In contrast, experts tended to skip steps by describing a small set of visual characteristics and then anchoring the descriptors to TCM patterns and patient symptoms. This is consistent with research related to medical cognition in which experts tend to use exemplars or are able to notice patterns earlier in the diagnostic process.[20]

Another notable pattern that was observed in the pre–case history was the use of higher-level comparator conceptual operators by experts earlier in the diagnostic process. One of the novices did not use any comparator conceptual operators for 2 of the tongues, and other novices tended to use comparators much later in the diagnostic process than experts. This is also consistent with the literature in which experts use exemplars to guide the problem-solving process. Experts draw on previous experiences to guide their reasoning processes. We also found that the experts in our study had more accurate diagnoses after reading the case history information. This may indicate that TCM practitioners are using domain knowledge and exemplars, processes consistent with Western clinicians.

Experts were more likely to describe the details of the tongue characteristics later in the sequence and tended to come to the diagnostic hypothesis later in the sequence than novices. This may be consistent with Crowley et al's work, in which experts demonstrated longer reasoning chains than novices.

Although the sample size was too small to effectively analyze for practice variation, we noted that 4 of the 7 of the practitioners identified 1 of the major diagnoses for each tongue. This

is consistent with the work by Hogeboom et al and Sherman in which multiple diagnoses were generated for each patient and in which inter-rater reliability was relatively low.[8,9] This is consistent with practice variation as seen in Western medical practice that is considered suboptimal because the variations can lead to inconsistent care delivery and subsequent unnecessary resource use.[53] Future work with a larger sample size, live patients, and examination of associated treatment plans and outcomes would shed more light on the practice variation in the TCM field.

Post–case history diagnostic reasoning differed slightly between novices and experts. Novices were more likely to use low-level explanation operators earlier in the sequence. In contrast, experts continued to explore the tongue photographs before they explained the TCM patterns. Although both groups were similarly likely to recognize discrepancies, it was notable that experts were more likely to revise their diagnoses in light of the discrepancy. This is consistent with our diagnostic accuracy evaluation in which we examined diagnostic accuracy by self-assessment and by gold standard. Self-assessment scores were relatively consistent between individuals, with the pre–case history scenario confidence score ranking lower than the post–case history score. Though all practitioners' diagnostic accuracies (by gold standard) were improved by the case history information, expert scores increased more than the novice scores. This may be consistent with the research related to expertise and to Norman's work in rapid diagnosis in dermatology in which experts used the information to come to more accurate diagnoses earlier.[43]

Another distinguishing pattern was in expression of higher-level meta-reasoning, whereby experts were more likely to use meta-reasoning operators earlier in the sequence than novices. This is consistent with recent work in nursing and medical decision making in which reflective practice may improve the accuracy of decision making.[54,55]

TCM tongue diagnosis is a complex and visually intensive process. Future work using live patients might facilitate a deeper understanding of best practices in the integration of domain knowledge and identification of specific patterns associated with TCM diagnoses.

Acknowledgments

REFERENCES

1. Riddle JW. Report of the New York Commission on Acupuncture. *Am J Chin Med (Gard City N Y)*. 1974;2(3):289-318.
2. White A, Ernst E. A brief history of acupuncture. *Rheumatology*. 2004;43(5):662-663.
3. Herman PM, Craig BM, Caspi O. Is complementary and alternative medicine (CAM) cost-effective? A systematic review. *BMC Complement Altern Med*. 2005 Jun 2;5:11.
4. Hulme C, Long AF. Square pegs and round holes? A review of economic evaluation in complementary and alternative medicine. *J Altern Complement Med*. 2005;11(1):179-188.
5. Ni H, Simile C, Hardy AM. Utilization of complementary and alternative medicine by United States adults: Results from the 1999 national health interview survey. *Med Care*. 2002;40(4):353-358.
6. Burke A, Upchurch DM, Dye C, Chyu L. Acupuncture use in the United States: Findings from the National Health Interview Survey. *J Altern Complement Med*. 2006;12(7):639-648.
7. Sherman KJ, Cherkin DC, Eisenberg DM, Erro J, Hrbek A, Deyo RA. The practice of acupuncture: who are the providers and what do they do? *Ann Fam Med*. 2005;3(2):151-158.
8. Sherman KJ, Cherkin DC, Deyo RA, et al. The diagnosis and treatment of chronic back pain by acupuncturists, chiropractors, and massage therapists. *Clin J Pain*. 2006;22(3):227-234.
9. Hogeboom CJ, Sherman KJ, Cherkin DC. Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists. *Complement Ther Med*. 2001;9(3):154-166.
10. Kalauokalani D, Sherman KJ, Cherkin DC. Acupuncture for chronic low back pain: diagnosis and treatment patterns among acupuncturists evaluating the same patient. *South Med J*. 2001;94(5):486-492.
11. Zell B, Hirata J, Marcus A, Ettinger B, Pressman A, Ettinger KM.. Diagnosis of symptomatic postmenopausal women by traditional Chinese medicine practitioners. *Menopause*. 2000;7(2):129-134.
12. Liangyue, D, Yijun, G, Shuhui, H, et al. *Chinese Acupuncture and Moxibustion*. Beijing: Foreign Languages Press. 1987.
13. Kirschbaum B. *Atlas of Chinese Tongue Diagnosis*. Seattle, WA: Eastland Press; 2000.
14. Kitson A. The state of the art and science of evidence-based nursing in UK and Europe. *Worldviews Evid Based Nurs*. 2004;1(1):6-8.
15. Maclocia G. *Tongue Diagnosis in Chinese Medicine*. Rev ed. Vista, CA: Eastland Press; 1995.
16. Straus SE. *Evidence Based Medicine: How to Practice and Teach EBM*. 3rd ed. New York, NY: Elsevier/Churchill Livingstone; 2005.
17. Chapman G, Epstein AS. Cognitive processes and biases in medical decision making. In: Chapman G, Sonnenberg F. eds. *Decision Making in Health Care*. Cambridge, UK: Cambridge University Press; 2000.
18. Popper K. *The Logic of Scientific Discovery*. New York: Basic Books; 1959.
19. Popper K. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York, NY: HarperCollins; 1963.
20. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39(4):418-427.
21. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002;324(7339):729-732.
22. Brinkhaus B, Witt CM, Jena S, et al. Interventions and physician characteristics in a randomized multicenter trial of acupuncture in patients with low-back pain. *J Altern Complement Med*. 2006;12(7):649-657.
23. Zhang GG, Lee WL, Lao L, Bausell B, Berman B, Handwerger B. The variability of TCM pattern diagnosis and herbal prescription on rheumatoid arthritis patients. *Altern Ther Health Med*. 2004;10(1):58-63.
24. Wisemann, N. *Fundamentals of Chinese Medicine*. Brookline, MA: Paradigm Publications. 1996.
25. Norman G. From theory to application and back again: Implications of research on medical expertise for psychological theory. *Can J Exp Psychol*. 2005;59(1):35-40.
26. Benner P. *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*. Reading, Massachusetts: Addison-Wesley; 1984.
27. Benner PE, Tanner CA, Chesla CA. *Expertise in Nursing: Caring, Clinical Judgment, and Ethics*. New York: Springer; 1996.
28. Buckingham CD, Adams A. Classifying clinical decision making: interpreting nursing intuition, heuristics and medical diagnosis. *J Adv Nurs*. 2000;32(4):990-998.
29. Ericsson KA, Lehmann AC. Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annu Rev Psychol*. 1996;47:273-305.
30. Ericsson KA, Smith J, eds. *Toward a General Theory of Expertise: Prospects and Limits*. Cambridge, UK: Cambridge University Press; 1991.
31. Kee F, Jenkins J, McIlwaine S, Patterson C, Harper S, Shields M. Fast and frugal models of clinical judgment in novice and expert physicians. *Med Decis Making*. 2003;23(4):293-300.
32. Sox HC, Marton KI, Higgins MC, Blatt MA. *Medical Decision Making*. Woburn, MA: Butterworth-Heinemann; 1988.
33. Hassebrock F, Prietula MJ. Protocol-based coding scheme for the analysis of medical reasoning. *Int J Man Mach Stud*. 1992;37(5):613-652.
34. Cioffi J. Heuristics, servants to intuition, in clinical decision-making. *J Adv Nurs*. 1997;26(1):203-208.
35. Krupinksi EA. Practical applications of perceptual research. In: Beutel J, Kudenl HL, Van Metter RL, eds. *Handbook of Medical Imaging. Vol 1: Physics and Psychophysics*. Bellingham, WA: SPIE Press; 2000:895-929.
36. McGuiness B, Axford R. Exploring nursing knowledge by using digital photography. *Stud Health Technol Inform*. 1997;46:281-287.
37. Robbins RM. The decubitus wound assessment database: An evaluation of student attitudes to the use of cognitive tools in an undergraduate nursing programme. Paper presented at: Proceedings of the International Medical Informatics Nursing Informatics working group; April 28-May 3, 2000; New Zealand.
38. Visco DC, Shalley T, Wren SJ, et al. Use of telehealth for chronic wound care: a case study. *J Wound Ostomy Continence Nurs*. 2001;28(2):89-95.
39. Doi K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Phys Med Biol*. 2006;51(13):R5-R27.
40. Krupinski EA. The importance of perception research in medical imaging. *Radiat Med*. 2000;18(6):329-334.
41. Kundel HL, Wright DJ. The influence of prior knowledge on visual search strategies during the viewing of chest radiographs. *Radiology*. 1969;93(2):315-320.
42. Tanner WP, Swets JA. A decision-making theory of visual detection. In: Yantis S, ed. *Visual Perception*. Philadelphia, PA: Psychology Press; 2001:48-59.
43. Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol*. 1989;125(8):1063-1068.
44. Leblanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic

hypothesis on the interpretation of clinical features. *Acad Med.* 2002;77(10 Suppl):S67-S69.

45. Crowley RS, Naus GJ, Stewart J, Friedman CP. Development of visual diagnostic expertise in pathology—an information-processing study. *J Am Med Inform Assoc.* 2003;10(1):39-51.
46. Crowley RS, Naus GJ, Friedman CP. Development of visual diagnostic expertise in pathology. *Proc AMIA Symp.* 2001:125-129.
47. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* New York, NY: Houghton Mifflin Company; 2002.
48. Dickson J, McLennan J, Omodei MM. Effects of concurrent verbalization on a time-critical, dynamic decision-making task. *J Gen Psychol.* 2000;127(2):217-229.
49. Henry SB, LeBreck DB, Holzemer WL. The effect of verbalization of cognitive processes on clinical decision making. *Res Nurs Health.* 1989;12(3):187-193.
50. Kuusela H, Paul P. A comparison of concurrent and retrospective verbal protocol analysis. *Am J Psychol.* 2000;113(3):387-405.
51. NVivo [computer program]. Version 2.0. Doncaster, Victoria, Australia: QSR International Pty Ltd 2002.
52. Newell A. *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall; 1972.
53. O'Neill L, Kuder J. Explaining variation in physician practice patterns and their propensities to recommend services. *Med Care Res Rev.* 2005;62(3):339-357.
54. Rew L. Acknowledging intuition in clinical decision making. *J Holist Nurs.* 2000;18(2):94-108, 109-113.
55. Rew L, Agor W, Emery MR, et al. Intuitive skills in crisis management. *Nursingconnections.* 2000;13(3):45-54.