

**Report to the Little Hoover Commission On
The Acupuncture Regulation Project:
Evaluation and Comparison of
California's License Examination and
NCCAOM's Certification Examinations**

Prepared For:

**Little Hoover Commission
925 L Street, Suite 805
Sacramento, CA 95814**

Prepared By:

**Lawrence S. Meyers. Ph.D.
Department of Psychology
California State University, Sacramento
Sacramento, CA 95819**

Under Independent Consultant Agreement C17849

June 20, 2004

Acknowledgements

I would like to acknowledge the efforts of several individuals who were involved in one or another capacity in this project. Stephen Klein of Rand Corporation played an important consultant role. He suggested the structure of some analyses, and he designed and interpreted the statistical analyses of the comparability of passing scores and language groups. Steve read a draft of this report and provided extremely valuable feedback and suggestions on several of technical issues. Deficiencies in the report, however, should still be attributable to me.

Elita Burmas was the research analyst who worked with me during this project. She performed all of the statistical analyses that we needed to do after entering relatively large quantities of data on the test items. Elita also took the lead on mapping the tasks identified in California's occupational analysis to the tasks documented in the national job analysis. Most of the tables we present in this report were put together by her as well.

This project called for extensive contact with the national and state testing programs. We deeply regret that NCCAOM experienced the tragic loss this year of its CEO, Christina Herlihy. While Christina was undergoing medical treatment during the winter Betsy Smith filled in overseeing this project and communicating with me. After Christina passed away in late March, Betsy and then Debra Persinger, who became NCCAOM's interim CEO, had to assume leadership on this project. Betsy and Debra were both very conscientious and gracious throughout this difficult period.

The individuals representing the California exam worked diligently to provide us with a considerable amount of information. Marilyn Nielsen, the Executive Officer of the Acupuncture Board, was very responsive to my various requests for information. Tracy Ferrel of the Office of Exam Resources came on as Chief a couple of years after the occupational analysis was completed. It is a mark of her competence that she was able to compile from the documentation in the files responses to some very specific questions and requests that I made concerning that analysis. Both Marilyn and Tracy consistently acted in the highest professional manner and were extremely cooperative.

Finally, Hattie Hanley of the Little Hoover Commission was the Project Manager for this endeavor. Her ability to see what was needed to make this evaluation a reality, her resourcefulness and support in immediately providing me with what I needed, and her insightful suggestions never ceased to impress me.

Table of Contents

<i>Acknowledgement.....</i>	<i>2</i>
<i>Purpose of the Comparison to be Made Here.....</i>	<i>5</i>
<i>The Two Test Developers.....</i>	<i>5</i>
<i>Abbreviated Calendar of Requests for Information to the Testing Programs.....</i>	<i>6</i>
<i>Summary of Material Reviewed.....</i>	<i>9</i>
<i>Double Test Taker Analysis.....</i>	<i>9</i>
<i>Some Preliminary Issues.....</i>	<i>10</i>
<i>Confidentiality Agreements.....</i>	<i>10</i>
<i>Licensing versus Certification.....</i>	<i>11</i>
<i>Some Context To Frame This Report.....</i>	<i>13</i>
<i>Goal of an Acupuncture License Exam.....</i>	<i>13</i>
<i>The Test Score as an Indicator of Qualifications.....</i>	<i>14</i>
<i>Testing As the Foundation for Inferences.....</i>	<i>14</i>
<i>Professional Standards Underlying Testing.....</i>	<i>14</i>
<i>Overview of a Valid Testing Process.....</i>	<i>15</i>
<i>Structure of the Present Report.....</i>	<i>16</i>
<i>Stage 1: Performing the Occupational Analysis.....</i>	<i>16</i>
<i>Formation of SME Panels.....</i>	<i>17</i>
<i>Developing the Job/Occupational Analysis Questionnaire.....</i>	<i>19</i>
<i>Organizing the Tasks and Knowledge Elements.....</i>	<i>25</i>
<i>Sampling Practitioner Respondents for the Job/Occupational Questionnaire.....</i>	<i>27</i>
<i>Distinguishing the Important Tasks and Knowledge Elements.....</i>	<i>34</i>
<i>Reliability of the Task and Knowledge Elements Ratings.....</i>	<i>36</i>
<i>Linking Tasks and Knowledge Elements.....</i>	<i>38</i>
<i>Stage 2: Examination Development.....</i>	<i>38</i>
<i>Weighting the Content Areas.....</i>	<i>38</i>
<i>Content of the Examinations.....</i>	<i>41</i>
<i>Mapping the California and National Content to Each Other.....</i>	<i>42</i>
<i>Crafting Test Items.....</i>	<i>47</i>
<i>Stage 3: Administration of the Exam.....</i>	<i>49</i>
<i>General Considerations.....</i>	<i>49</i>
<i>The Issue of Testing In Multiple Languages.....</i>	<i>49</i>
<i>Stage 4: Quality Assessment.....</i>	<i>54</i>
<i>The Context For Performing a Quality Assessment.....</i>	<i>54</i>
<i>The Structure Of This Section Of the Report.....</i>	<i>55</i>
<i>Central Tendency Statistics For the California and NCCAOM Samples as a Whole.....</i>	<i>55</i>
<i>Central Tendency Statistics For the California Language Groups.....</i>	<i>57</i>
<i>Variability Statistics For the California and NCCAOM Samples as a Whole.....</i>	<i>57</i>
<i>Reliability Statistics For the California and NCCAOM Samples as a Whole.....</i>	<i>58</i>
<i>Reliability For the Different Language Groups in the California Examinations.....</i>	<i>60</i>
<i>Item Difficulty Statistics For the California and NCCAOM Samples as a Whole.....</i>	<i>65</i>
<i>Point-Biserial Statistics For the California and NCCAOM Samples as a Whole.....</i>	<i>69</i>

Table of Contents (Continued)

<i>Stage 5: Pass Point Analysis</i>	74
<i>Establishing a Pass Point</i>	74
<i>Passing Rates</i>	76
<i>The Issue of “Must Pass” Requirements</i>	79
<i>Evaluation Summary</i>	80
<i>Some Consequences of Choosing One Test or the Other</i>	82
<i>The Choice Between the California and National Examination Programs</i>	83
Attachments	
<i>Attachment A</i>	85
<i>Attachment B</i>	89
<i>Attachment C</i>	95
<i>Attachment D</i>	99
<i>Attachment E</i>	104
<i>Attachment F</i>	106
<i>Attachment G</i>	109
<i>Attachment H</i>	115
<i>Attachment I</i>	122
<i>Report Supplement</i>	124

Report to the Little Hoover Commission On The Acupuncture Regulation Project: Evaluation and Comparison of California's License Examination and NCCAOM's Certification Examinations

**Lawrence S. Meyers, Ph.D.
June 2004**

Purpose of The Comparison To Be Made Here

In September of 2002, the California State Legislature through Senate Bill 1951 and Assembly Bill 1943 asked the Little Hoover Commission to take under consideration several issues pertaining to the licensing of those professionals practicing Acupuncture and Oriental Medicine in the State of California. The particular issue addressed by the present report is contained in Section 4934.1 (a) (3) of Senate Bill 1951, Chapter 714 which reads as follows:

Evaluate the national examination, administered by the National Certification Commission for Acupuncture and Oriental Medicine, and make recommendations as to whether or not the national examination should be offered in California in lieu of, or as part of, the state examination.

The state has developed the California Acupuncture License Exam (CALE) to determine who possesses the minimal level of knowledge to be licensed to practice here. The National Certification Commission for Acupuncture and Oriental Medicine (NCCAOM) offers five separate modules that test for similar content. Based on the above charge from the Legislature, this evaluation took the following form: To the extent that it was possible, the examination development and testing processes used by the State and NCCAOM were matched with each other so that a comparison and evaluation could be made.

The Two Test Developers

As a general principle, the groups that develop licensing or certification examinations are composed of testing professionals who presumably have training and expertise in job analysis, test development, test validity, test assessment, and psychometrics. In addition, they also possess a reasonable degree of statistical sophistication so that they can perform the

necessary analyses of the testing data. Typically, those testing and quantitative technical skills are not possessed by practicing acupuncturists, members of licensing or certification boards, or program administrators.

In this state the Office of Examination Resources (OER) of the California Department of Consumer Affairs has been given the responsibility by the Legislature to create the license exam for acupuncture. When a reference is made to the developers of the California test, it is OER that is being referenced. At the national level, NCCAOM, which is a private sector non-profit organization, administers a certification program that, according to their website, establishes “standards of competence and safety.” But NCCAOM contracts with Applied Measurement Professionals, a private sector consulting firm that specializes in developing credentialing examinations, for all of the test development and technical work associated with the testing process. Each time some technical information was requested of NCCAOM, it wrote a contract with and paid Applied Measurement Professionals to supply us with certain data or to conduct and pass along certain statistical analyses. Although this report will refer regularly to NCCAOM, it should be remembered that all of the test development work, including conducting the job analysis, developing and administering the test, and analyzing the test results were actually done by their technical consulting firm.

Abbreviated Calendar of Requests for Information to the Testing Programs

Table 1a through Table 1c present a rough timeline of the more salient communications that have transpired in this report period in order to acquire the information necessary to compare the two testing programs. The communication exchange for one analysis, dubbed the double test taker analysis, is not covered in this time line since this analysis was not possible to conduct. A brief description of what that analysis was intended to be is presented in the next section.

As can be seen from these timelines, the initial request for information was made in the middle of September of 2003. OER was very prompt in their response to the requested information. They produced quite a bit of material in about two months or so. When it was determined that the information they provided was insufficient in certain areas and further materials were requested, OER made available further materials in a timely fashion.

As can be seen in the three tables below, the response from NCCAOM did not arrive for six months. It was delayed, they indicated, because they were in the process of revising their testing procedure and were unable to provide the information I requested any earlier. However, when it finally arrived six months following my initial request, the response was not forthcoming. Most of the information that was sent was already in my possession, that is, they were excerpts of NCCAOM’s job analysis that was publicly available on their website. Further, in response to my requests for statistical information, they repeatedly stated that it was not available. Finally, in response to my further appeal asking that the materials be provided to me by April 30, some of the information I had initially requested was provided in a mailing I received on the day of the deadline, just two months prior to the time that my final report was to be turned in to the Little Hoover Commission.

Table 1a
Calendar of Information Requests and Responses
September 2003 to End of Calendar Year

Date	Nature of Communication
September 19, 2003	Initial request for information was sent out to the State and to NCCAOM. This is included as Attachment A.
Late September (est.)	Received copy of job analysis downloaded from the NCCAOM website.
October 17, 2003	Met with Marilyn Nielsen, the Executive Officer of the Acupuncture Board and Dr. Tracy Ferrel of OER to speak about my request. At that time, I signed their standard confidentiality agreement.
November 26, 2003	Received a large set of material from Tracy Ferrel of the OER of the California Department of Consumer Affairs including occupational analysis, item and test statistics, and pass point setting procedure.
December 3, 2003	I raised further statistical questions with the State.
December 11, 2003	Response from Tracy Ferrel of OER
Mid December (est.)	Phone call from Betsy Smith of NCCAOM indicating they plan to start processing my request for information on January 12, 2004

Table 1b
Calendar of Information Requests and Responses
January 2004 Through March 2004

Date	Nature of Communication
January 14, 2004	Sent extensive list of questions regarding the State's documentation to Tracey Ferrel of OER and Marilyn Nielsen following up on material sent at end of November. This is included as Attachment B.
February 1, 2004	Email to Betsy Smith of NCCAOM asking when they expect to have the requested information to me.
February 5, 2004	Betsy Smith answered my inquiry indicating NCCAOM would mail material late the following week.
March 5, 2004	Received a packet of material from NCCAOM containing responses to my inquiry of September 19, 2003. Many of the narrative responses were copied and pasted from their job analysis. In response to my statistical questions, the response was repeatedly "not available." Relevant responses to these statistical inquiries were finally sent by NCCAOM at the very end of April.
March 8, 2004	Email to Betsy Smith indicating that I believed that at least some of the information I had requested, such as standard item analysis, was available despite their indicating "not available."
March 10, 2004	Received email from Christina Herlihy of NCCAOM telling me not to worry because they do conduct standard item analyses but declining to send that material as "it will incur a huge financial expense."
March 10, 2004	Responded to Christina Herlihy of NCCAOM indicating that I would be happy to receive photocopies of the item analyses that are most likely to be in their files already as part of the test documentation and that there was no need to reproduce the analyses for me at this time.
March 12, 2004	Received response from Tracy Ferrel of OER to my January 14, 2004 inquiry.
March 28, 2004	Sent letter to Betsy Smith of NCCAOM summarizing the information that I needed to complete the report but which they have not yet sent. Gave a deadline of April 30, 2004 for receipt. This is included as Attachment C.
March 28, 2004	Sent letter to Marilyn Nielsen of the Acupuncture Board and Tracy Ferrel of OER summarizing the information that I needed to complete the report but which they have not yet sent. Gave a deadline of April 30, 2004 for receipt. This is included as Attachment D.

Table 1c
Calendar of Information Requests and Responses
April 2004 Through June 2004

Date	Nature of Communication
April 1, 2004	Received material back from Tracy Ferrel of OER responding to my inquiry of March 28, 2004.
April 16 and 23, 2004	Betsy Smith of NCCAOM left telephone messages saying that NCCAOM is working on providing me with some of the information I requested and that I will hear from their attorney in the near future.
April 18, 2004 to April 28, 2004	Exchanged communications with Jeff Glassie of Shaw Pittman, the attorney for NCCAOM, discussing the confidentiality agreement that they wished me to sign.
April 29, 2004	After some modifications to the confidentiality agreement prepared by NCCAOM's attorney, I signed it.
April 30, 2004	Received materials from NCCAOM in response to my March 28 letter.

Summary of Material Reviewed

Both OER and NCCAOM provided their most recent occupational and job analyses, which carried dates of 2001 and 2003, respectively. Supplemental information that I requested when these reports proved incomplete was also sent by these agencies. In addition, a considerable amount of computerized statistical output was provided by OER and NCCAOM. This included item and test analyses as well as passing rate information. The Little Hoover Commission was kind enough to contribute other documents when it could, such as the transcript of a discussion that took place between the Acupuncture Board and Christina Herlihy who was CEO of NCCAOM at the time. A small amount of information was also gleaned from the websites of the Acupuncture Board and NCCAOM. Sources for some of the more specific pieces of information are cited in the report in the context of the particular topics covered.

Double Test Taker Analysis

As a smaller part of the work that was to be done for this report, we had hoped to conduct an analysis on the examination scores for those individuals who had taken both the CALE and national exams; this was the basis for naming this the “double test taker analysis.” The intent was to compare the average score on the State to the national exam to contribute to the estimation of which might be “easier” (with the recognition that difficulty and quality may be quite different) as well as to correlate these scores to determine the extent to which they seemed to measure a common set of constructs.

To perform such an analysis, it is necessary to match individuals from both exams. That is, we must know Candidate A's score on the CALE and Candidate A's score on the national exam. Matching must be done in such a way as to minimize the chances of mistaken

identities. For example, it is possible for two or more individuals to have the same name. The more information we have about individuals for each exam, the better the quality of the match.

At considerable effort, the Acupuncture Board supplied us with a confidential list of individuals who had taken their licensing examination. This list contained several identifiers, none of which we had intended to record in our data file once we obtained the matching score from the national exam but which could be used in the process of matching these persons to the national list. However, NCCAOM referred to its policy of not releasing candidate names and was therefore unable to comply with our request.

We attempted to negotiate various solutions to this problem, all of which failed. For example, we considered sending the list of names and identifiers supplied by the Acupuncture Board to NCCAOM for them to go through their own files and find the matches and send us the data. Many of the procedures that we offered are actually used relatively frequently, but on advice of their attorney the Board felt that it could not release the names or cause the names to be released to NCCAOM. In addition, NCCAOM offered no workable solution to the problem.

In a last attempt to resolve the issue, I spoke with the attorney representing NCCAOM, Jeff Glassie of Shaw Pittman, on April 9, 2004 and he indicated with finality that NCCAOM could not comply with my request. He said he would send something in writing. His e-mail note indicating that NCCAOM would not send the material is dated April 16, 2004 and may be seen as Attachment E. Without the NCCAOM data, this double test taker analysis was unable to be launched.

As somewhat of a balance to not receiving this information, at least some aspects of the information we were hoping to glean from this analysis can also be approached from a somewhat different direction, and this project was structured to explore that approach. It is possible to speak to difficulty levels of the exams, at least indirectly, by comparing the passing percentages of candidates on each exam and examining the item difficulty of the test questions. As for measuring common constructs, we performed an assessment of the tasks from each job/occupational analysis to study the degree to which the tests targeted similar content. The results of these analyses will be presented in the report that follows.

Some Preliminary Issues

Confidentiality Agreements

During the meeting with Dr. Ferrel of OER and Marilyn Nielsen, the Executive Officer of the Acupuncture Board, I signed a relatively standard Examination Security Agreement. Essentially, it restricted me from giving certain secured information, such as the examination items and correct answers as well as any information identifying individual test takers, to anyone. Upon my signing the agreement, both OER and the Acupuncture Board started to supply me with the information that I had requested so that I could use that material in my report to the Little Hoover Commission. Neither OER nor the Acupuncture Board ever questioned or attempted to negotiate what would be contained in my report, that is, they adopted a position of providing what they could and let the evaluation process take its course.

On April 29, 2004 and following much negotiation I signed a confidentiality agreement with NCCAOM. Perhaps partly because it is a private sector organization, perhaps partly because they contract for their test development and test analysis with a private sector consulting firm, perhaps partly because of a variety of other reasons, NCCAOM was particularly concerned with what information of the material they were about to turn over to me might be contained in my report to the Little Hoover Commission. They believed, properly, that once turned over to the Commission, the content of that report was in the public domain and outside of my and, more importantly to them, their control.

At any rate, NCCAOM was sufficiently concerned about the release of information they considered to be potentially sensitive regarding their tests that the agreement contained, from their perspective, certain safeguards. The following safeguard was the topic of much discussion and some revision before I agreed to it:

As general and overriding obligations and limitations, I understand that I may review and analyze the information provided by NCCAOM and report my findings solely to the above named organizations [my direct contractor UCSF Center or the Health Professions and the Little Hoover Commission], but that I am not authorized to report, reveal, or disclose—without the prior written approval of NCCAOM—information provided by NCCAOM...

My concern, expressed to them in clear terms, was that abiding by the terms of the agreement might give the appearance that NCCAOM was in some way controlling, editing, or censoring what went into this report. It was agreed that if NCCAOM did refuse to authorize any information being placed in this report, their refusal would be made in a public manner to both me and to the staff of the Little Hoover Commission. Let it be known here that I indicated three times that I would be using portions of the information sent to me on April 29 (received April 30) and that in all three instances NCCAOM offered no objection to their use.

Licensing versus Certification

One can envision a continuum of marks of distinction that recognize the expertise demonstrated by individuals engaged in designated activities. Licenses to practice a profession occupy one end of that scale and certifications occupy the other end.

At least in their somewhat extreme situations, there are noticeable differences between a licensing process and a certification process. Licenses are issued by a governmental agency, usually a state. They allow individuals to perform some actions or provide some services that in lieu of a license would be prohibited. A license is usually required when issues of public health or safety are in question. For example, one cannot legally drive a motorized vehicle on a public road in California without a driver's license, and one cannot practice acupuncture in California without an acupuncture license.

Certifications are usually issued by private sector organizations. Awarding a certification to an individual is typically contingent on that person having accomplished one or more stated objectives such as having passed an examination or having satisfactorily completed a program of study. Certification may increase the competitive advantage of its holder in the

marketplace, but it is not ordinarily a legal requirement in order to be allowed to provide some service.

Between these two extremes lies an area in which certifications carry a more authoritative voice, blurring the line sometimes between the license and the certification. For example, several respected professions not licensed by the state will still require its practitioners to be certified in order to publicly proclaim themselves as members of that profession. For example, a Certified Public Accountant must be certified. Teachers, school psychologists, and school counselors represent some other professions for which certification is required.

Sometimes certification is used to identify those members of the profession who have achieved or demonstrated designated knowledge or skills beyond the generalist. To be Board Certified in Psychiatry indicates that those licensed physicians have been recognized as specialists by the medical profession. This report will limit itself to the differences and similarities between licensure and certification in their less ambiguous states.

It is possible for a licensure system to interface with a certification system. Such interfacing can be done by either using or requiring such certification in the licensing process, a policy that is made by individual states. In the area of acupuncture and oriental medicine, this interfacing has been achieved by about a dozen and a half states that use NCCAOM's exams in their licensing procedures.

Despite the differences between licensing and certification, at least in terms of the examination processes that have developed in the particular instances under study here, the licensure testing process of the State of California and the certification exam process of NCCAOM seem more alike than different. Here is the mission statement of the Acupuncture Board:

The mission of the Acupuncture Board is to protect and educate the public through appropriate regulation of licensure, education standards, and enforcement of the Acupuncture Licensure Act, which includes Oriental Medicine.

While a bit different because it is not a state agency charged with certain public protections, the mission statement of NCCAOM is similar to that of the Acupuncture Board:

The mission of the National Certification Commission for Acupuncture and Oriental medicine (NCCAOM) is to establish, assess, and promote recognized standards of competence and safety in acupuncture and Oriental medicine for the protection and benefit of the public.

If the differences between the two agencies are relatively similar, the differences between their two examination processes appear to be minimal at best. Both California and NCCAOM engaged in a comprehensive job analysis for the geographic regions currently served by their testing programs. Both have developed their examinations based on that work. Both maintain relatively large item banks and replenish them regularly. Both perform similar types of statistical analyses on their test results. And both establish passing criteria using virtually the same methodology.

I am aware that it has been questioned whether NCCAOM's certification examination process can logically be transported over to California and used as a licensing exam process

here. My answer to this concern is “yes” for two reasons. First, and of lesser importance, the national exam appears to be used by more than a dozen states, such as Arizona, Colorado, Florida, and Oregon, in their licensing process. Thus, several other states have decided that a certification exam can serve licensing purposes. Second, and much more to the point, regardless of what it is called, the NCCAOM examination modules taken together very much resemble the exam used in California; that is, if one does not dwell on its label, the national testing process certainly was developed in a similar manner to the CALE and it very much looks like the CALE. Functionally, then, it should not be excluded from consideration at the start of this evaluation process but must be regarded as potentially being able to potentially serve as a licensing exam in California.

In summary, in my opinion an initial challenge to the viability of the comparison of the CALE and national exams does not carry sufficient weight to stop the review, evaluation, and comparison process at this point. Thus, the comparison and evaluation of the two exams for future use as a licensing test in California is the subject of the remaining portion of this report.

Some Context To Frame This Report

Goal of an Acupuncture License Exam

In order to provide a context for comparing the two tests, it is necessary to keep in mind the goal that is to be served by the license exam. The intent of any licensing process is to protect the health, safety, and welfare of the public by restricting practice of certain professions to only those who are qualified. At the same time, licensing is intended to be nonrestrictive. That is, while the standards may be strict, they should “...not be so stringent as to unduly restrain the right of qualified individuals to offer their services to the public” (Standards for Educational and Psychological Testing, 1999, p. 156).

The trade-off between protecting the public and minimizing the restrictiveness of licensure is handled in the acupuncture situation in a very traditional manner. Consistent with other licenses awarded by the State of California, and consistent with the licensing practices of most other states, the license exam for acupuncture is intended to identify “...the minimal acceptable level of competence required in California for the safe and effective practice of acupuncture” (Acupuncture Board website, April 4, 2004). The goal of the license examination is to identify those licensing candidates who meet this criterion.

The Test Score as an Indicator of Qualifications

On the basis of a candidate's score on the examination, the Acupuncture Board determines whether or not that candidate has demonstrated the minimal level of competence to be awarded a license to practice acupuncture. Such an inference from the test results—that by earning a specified score candidates have shown themselves to be minimally competent—must rest upon a solid, valid, and reliable foundation for the State of California to assert that it has properly assessed the qualifications of applicants. That is, the test score can be used to make an inference about candidates' competency if and only if the entire process that produced it—the entire test development, administration, and pass point decision process—is sufficiently strong (valid) to justify the inference.

Testing As the Foundation for Inferences

When evaluating a test for use in a particular application, such as a test that is used to measure minimal competency of applicants seeking a license to practice acupuncture, it is the validity and reliability of the foundation for the test that is carefully scrutinized. This foundation comprises an extensive test development process that starts with a comprehensive description of the profession, goes through the construction, administration, and quality assessment of the examination, and ends with a determination of the passing score as the lower bound of minimal competency. The review process performed here in response to Section 4934.1 (a) (3) evaluates this test development process for both the CALE and the NCCAOM examinations.

Professional Standards Underlying Testing

Testing practice has evolved to the point where there is relative agreement on the general process that needs to be in place to support the inferences from test scores that agencies wish to make. These standards are documented in publications and professional presentations and are periodically revised to incorporate recent advances in the field. Sources include professional publications, professional books, government and agency documents, material disseminated at professional meetings, and even academic textbooks in testing. The following list covers some of those resources used to structure this present evaluation. I have also included in the list some documents supplied by both OER and NCCAOM since these documents are consistent with the professional standards articulated by the more public and nationally recognized sources.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D. C.: Author.

- Anastasi, A., & Urbana, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Holt, Rinehart and Winston.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43(166), 38290–38309.
- Gion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hamilton, K. (2000). *Examination validation policy*. Sacramento, CA: California Department of Consumer Affairs.
- Meyers, L. S. (2004, March). *Searching for validity evidence*. Presented at the 18th Annual Spring Conference of Personnel Testing Council of Northern California. Emeryville, CA.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed) New York: McGraw-Hill.
- National Commission for Certifying Agencies. (2003). *Standards for the accreditation of certification programs*. Washington, D. C.: National Organization for Competency Assurance.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green, OH: Author.

Overview of a Valid Testing Process

Allowing for some minor variation, a generally accepted testing process can be outlined based on the professional standards currently in place. Applied to acupuncture licensing, the process entails properly completing the following steps:

- *Perform a job or occupational analysis.* This identifies the important tasks that are performed by licensed practitioners from which the knowledge, skills, and abilities (KSAs) underlying those tasks can be identified. Although its goals are relatively straightforward, it is a complex, time consuming, and professionally demanding work product that serves as the foundation of the validity evidence supporting the use of a license examination.
- *Examination development.* The important KSAs are mapped on to an examination plan that is designed to test them by appropriate formats. In the present situation, the single format used is a written objectively scored test. Test items are prepared based on commonly accepted principles of item writing.
- *Administration of the examination.* Every candidate is administered the same or comparable examination in the same manner. This enables the exam scores of candidates to be compared to each other and eventually to an established pass point.

- *Quality assessment of the examination.* The inferences about competency level of the candidates based on their test scores assumes that the examination was of sufficient quality to warrant such inferences being drawn. Exam quality is assessed on a variety of points that can be statistically indexed. These include (a) performance of individual test items regarding, for example, item difficulty and item-total (point-biserial) correlations; (b) demonstration of comparability of different forms such as those used across exam administrations or those used for different languages within a single exam administration as demonstrated by equivalence analysis; (c) the reliability of the test scores in terms of an index of internal consistency.
- *Determination of the pass point.* This is a structured process, often using some variant of what is known as an Angoff procedure, that attempts to identify the level of performance on the test corresponding to minimal competency.

Structure of the Present Report

The present report traces the development of each licensing exam by discussing the stages of test development outlined above: performing a job or occupational analysis, developing the examination development, administering the examination, assessing the quality of the examination, and determining the pass point for the exam. Each major phase of the testing process will be discussed in turn. Within each, the process followed by the State and national testing programs will be compared to the extent that the parties provided the information that was requested. Further, because both testing programs have recently been restructured, roughly only the last two to three years for each is considered. This time period, however, establishes the foundation for the two testing programs that will be existent over the next several years and is therefore the most relevant time period to consider in responding to the Legislative request. Evaluation opinions will be offered on an ongoing basis.

Stage I: Performing the Occupational Analysis

An occupational analysis is a comprehensive description of professional practice of a given occupation or profession. The analysis must be sufficient to allow for the development of a license examination. There are certain steps that must be accomplished by any appropriately conducted occupational analysis, and these are discussed in the sections that follow.

One of the first major goals to be achieved by an occupational analysis is to identify the important tasks that are performed by practitioners. But because the test developers are rarely practitioners of the occupation under study and should almost never act in that capacity in any case, they are not in position to put forward the tasks that practitioners ordinarily perform. They will therefore poll a sample of the practitioner population about these tasks by having these practitioner rate elements associated with the tasks.

In order to put the tasks before the sample, they need to generate the tasks in the first place. To achieve the goal of identifying these tasks, test developers recruit knowledgeable informants representing the profession. These informants, generally called subject matter

experts (SMEs), are placed on panels and work directly with the test developers in generating an initial list of tasks and KSAs.

Formation of SME Panels

SMEs serve as working liaisons to the test developers. They inform the test developers about the profession and are put to work by the test developers to create the initial list of tasks that practitioners perform in the ordinary and extraordinary conduct of the occupation. Members do not always need to meet face to face although it is often useful for that to happen from time to time.

California SME Panels: Description

As indicated in the occupational analysis (Occupational Practice, State of California, Department of Consumer Affairs, November 2001) the first part of this test developer-SME interaction process involved semi-structured onsite and telephone interviews with licensed acupuncturists. This group did not meet together and really did not need to do so. After organizing this information, the tasks were reviewed and edited by two additional SME panels. It is not clear if these latter two panels met face to face but it is likely.

For the purposes of this report, it is important to determine whether the SMEs participating in these panels were appropriate representatives of their profession. The official occupational analysis is silent on the makeup of these panels, but in response to our inquiries, Dr. Ferrel of OER in her March 10, 2004 response provided some much needed detail. The first panel contained 25 SMEs while the other two contained 8 SMEs. These counts are acceptable.

As to how well they represented the population of practitioners, it appears that both Northern and Southern California were appropriately represented in all panels. The first panel contained practitioners who were licensed 5 years or less as well as those who were licensed for 6 to 20 years; the last two panels contained only those who were licensed for the longer period of time. This is also acceptable.

No other demographic information concerning the SMEs was collected. OER interprets Government Code section 1870.2 and Civil Code section 1798.14 to prohibit them from collecting information “that is not necessary to fulfill their regulatory duties (e.g., gender, race, and national origin)” (Dr. Ferrel March 10 letter, p. 1). Nonetheless, OER appears to be sensitive to gender and ethnicity variables: “OER can assure you that the number of men and women included in the occupational analysis was relatively equal. In addition, because the profession is comprised of practitioners from several different cultures, an effort was made to interview practitioners from the major cultures represented (i.e., Chinese, Korean, and English)” (Dr. Ferrel March 10 letter, p. 2).

California SME Panels: Evaluation

The panels used by OER and the job that these panels were ask to accomplish are consistent with what is considered to be “best practice.”

California SME Panels: Recommendations

Since OER appears to be assembling panels exactly as they should, it is too bad that they do not appear to be able to collect some very important demographic information on their panelists. Precisely because the Chinese and Korean (and possibly even the Japanese) cultures play so conspicuous a role in acupuncture, it would appear that such information actually is necessary to be collected in order for OER to fulfill its regulatory duty. That is, documenting that all of the relevant cultures are represented in the development of this licensing exam speaks in part to the exam’s validity, and the State of California probably legally bound to administer valid licensing exams (i.e., produce evidence of validity to support using the exam as a basis for licensing acupuncturists). It is therefore recommended that such data collection be enabled.

NCCAOM SME Panel: Description

In contrast to California, NCCAOM assembled a single panel of 15 practitioners. NCCAOM’s job analysis (A National Job Analysis: Acupuncture and Oriental Medicine Profession Conducted for the National Certification Commission for Acupuncture and Oriental Medicine by Applied Measurement Professionals, Inc., August 2003) states that their panel “was reflective of the AOM profession in all relevant respects, for example: geographic, professional area and level of responsibility, educational background, gender, and work setting” (p. 3). Based on the names of the panelists, it is my guess that a range of cultures is probably represented as well.

NCCAOM SME Panel: Evaluation

While there are some advantages to using multiple smaller groups (e.g., each participant can have a greater voice, “group-think” at least in one direction is contained to a single group) using a single relatively large expert panel as an advisory group for the whole job analysis is certainly acceptable and may have its own advantages (e.g., more different viewpoints being expressed, the flexibility to form subcommittees).

Developing the Job/Occupational Analysis Questionnaire

One of the main functions performed by the SME panels is to serve as a resource in the development of a questionnaire that is to be administered to a sample of the practitioners in the geographic area in which the license or certification examination is to be used. This questionnaire is often labeled as a “Job Analysis Questionnaire,” an “Occupational Analysis Questionnaire,” or something similar.

The questionnaire serves as the cornerstone of the license exam because the responses to its questions will define the content of the profession. If the examination is to be used to determine if individuals are qualified to practice acupuncture, the licensing test must ask candidates occupationally relevant questions. This strategy to achieve validity (justifying the inference that the license exam score is an index of candidates’ level of competence) is known as content validation (i.e., the content of the occupation matching the content of the test is a source of validity evidence). The content validation strategy must be used in license examination development, and both OER and NCCAOM indeed pursued this route.

Scope Covered in the Questionnaire

Ordinarily, two domains are included in an analysis of a profession, the tasks that practitioners perform in their practice and the competencies that underlie the successful performance of these tasks. Competencies are thought of as the knowledge behind the tasks, as well as the abilities and skills needed in safely and competently carrying out the tasks.

Knowledge is most commonly tested by paper-and-pencil tests, such as multiple choice examinations. Abilities can also be tested by paper-and-pencil tests but are also commonly assessed by other types of exercises such as responses to interview or other open-ended questions, interacting in a role play exercise, evaluating a scenario that is presented, and so on. Skills are usually assessed by asking candidates to engage in behavior akin to what they would do in professional practice. Partly because California uses the term and partly because it probably comprises the bulk of what the CALE and the NCCAOM modules focus on, we will use the term “knowledge elements” to represent the competencies underlying the tasks.

Purpose of the Questionnaire

Determining the important tasks performed by practitioners and their related knowledge requirements serves at least two purposes in an occupational analysis. One purpose served by determining the important tasks and their related knowledge requirements is to define the parameters for performance testing. For example, in the California occupational analysis under the content area of “Providing Acupuncture Treatment” we find Task 112: “Locate points by implementing anatomical landmarks and proportional measurements to identify area for needle insertion.” In a performance test, candidates might very well be asked to physically locate particular points on a dummy or a volunteer. Since neither the CALE nor

the NCCAOM exams are hand-on performance tests, this purpose of determining the important tasks is not played out.

A second purpose served by determining the important tasks is that the knowledge underlying it can be specified and then tested in a multiple choice examination format. For example, two of the knowledge elements linked to Task 112 and its kin are Knowledge Element 180: “Knowledge of the physical landmarks and gross unit proportions used in point location,” and Knowledge Element 181: “Knowledge of the anatomical locations of points.” These knowledge elements can then be structured into test questions to determine if candidates know this information. Both the California and NCCAOM tests are paper-and-pencil and so this purpose of identifying the tasks and their related knowledge elements is (i.e., should be) fulfilled.

Identifying the Full Set of Tasks for Inclusion in the Job/Occupational Analysis Questionnaire

Task descriptions were generated as a foundation for both the CALE and the NCCAOM exams by panels of SMEs. These tasks then appeared on the occupational or job analysis questionnaire administered to a sample of practitioners. The process each developer used will be briefly discussed in turn.

OER’s Occupational Analysis

Although the occupational analysis produced by OER (before Dr. Ferrel was brought on as Chief of that unit) was incompletely documented on this matter, Dr. Ferrel’s March 10, 2004 memo supplied the missing information on which this portion of the report is based. The tasks were compiled from what I believe to be in-depth interviews with 25 practitioners whose composition appears to be representative of practitioners throughout California. This tabulation was reviewed independently by two additional SME panels of eight practitioners each to make sure that the list of tasks was complete. In all, a total of 163 tasks appeared on the questionnaire.

Evaluation of OER’s Work

In my opinion, this represents appropriate and high quality work consistent with professional standards. The occupational questionnaire could have asked respondents to provide any tasks that were missing or to evaluate the completeness of the list, but the process used to generate the list was adequate.

NCCAOM's Job Analysis

NCCAOM assembled what it called a "Job Analysis Advisory Panel" to oversee the entire process of the job analysis. This panel consisted of 15 practitioners who appeared to be representative of the geographic domain in which the test was to be applied. Tasks were compiled by this group of SMEs. In addition, a series of items was contained on the questionnaire asking for an assessment of the completeness of the list, and almost all of the respondents endorsed the list as they saw it. In all, a total of 284 tasks appeared on the questionnaire.

Evaluation of NCCAOM's Work

In my opinion, this represents work of high quality and certainly meets professional testing standards.

Identifying the Full Set of Knowledge Elements for the Job/Occupational Questionnaire

Knowledge elements are ordinarily derived from the tasks since, in paper-and-pencil tests it is these elements, rather than the tasks themselves, that are used as the basis for developing the test items. We will cover how each test developer dealt with this issue.

OER's Occupational Analysis

OER determined the knowledge elements associated with the tasks at the same time that the tasks were being documented. That is, when the first panel of 25 SMEs were asked to identify what tasks were performed by acupuncturists, these SMEs were also asked to indicate the associated knowledge elements that were "...required to perform those tasks in a safe and competent manner" (Occupational Practice, State of California, Department of Consumer Affairs, November 2001, p. 3). These knowledge elements also appeared on the occupational analysis questionnaire in addition to the tasks that we have already discussed. In all, a total of 269 knowledge elements appeared on the questionnaire.

Evaluation of OER's Work

In my opinion, the determination of knowledge elements and the way in which it was done is fully in accord with the standards of professional practice.

NCCAOM's Job Analysis

There is no mention in NCCAOM's job analysis report (A National Job Analysis: Acupuncture and Oriental Medicine Profession Conducted for the National Certification Commission for Acupuncture and Oriental Medicine by Applied Measurement Professionals, Inc., August 2003) of the knowledge elements that are associated with the tasks. That is, none appeared on the questionnaire or were documented as serving as the basis of examination development. Thus, the job analysis questionnaire asked for ratings of only the 284 tasks.

The absence of a list of knowledge elements is surprising and appears to be in conflict with test development standards in general and with the particular standards to which NCCAOM subscribes. As for the standards in general, the *Principles for the Validation and Use of Personnel Selection Procedures* on page 10 under the heading "Purposes for Conducting an Analysis of Work" states that:

One purpose is to develop selection procedures. Part of this process is an analysis of work that identifies worker requirements including a description of the general level of ability, skill, knowledge, or other characteristics needed.

NCCAOM's own internal standards also address this issue. The *Standards for the Accreditation of Certification Programs*, produced by the National Commission for Certifying Agencies, presents several standards under the topic of Assessment Procedures. Standard 10 reads as follows:

The certificate program must analyze, define, and publish performance domains and tasks related to the purposes of the credential, and the knowledge and/or skill associated with the performance domains and tasks, and use them to develop specifications for the assessment instruments.

Under the Commentary for Standard 10, the importance of knowledge elements is mentioned several times. Commentary B, for example, reads in part:

Validation of performance domains, tasks, and associated knowledge and/or skills is typically accomplished by conducting a survey of current certificants and/or individuals providing services or performing a job consistent with the purpose of the credential...

Commentary C reads in part:

Analysis of ratings information collected in the survey should determine how and to what degree the performance domains, tasks, and associated knowledge and/or skills relate to the purpose of the credential...

Commentary D reads in part:

A table of specifications should be prepared for each assessment instrument specifying the weighting of performance domains, tasks, and associated knowledge and/or skills to be included. The weighting system should be based primarily on data collected from survey participants, with informed review and interpretation provided by a panel of subject-matter experts. Decision rules used to eliminate performance domains, tasks, and associated knowledge and/or skills from the specification table should be explained...

And, as a last example, Commentary E reads in part:

Because rapid changes may occur in knowledge and/or skills and in technology, it is important that certification programs periodically review performance domains, tasks, and associated knowledge and/or skills in the specifications to ensure that they are current...

It seems quite clear from the above excerpts of the Certification Standards that knowledge elements (in this case because skills would not ordinarily be measured in the paper-and-pencil exam that NCCAOM was planning, we would not expect them to necessarily focus on identifying the skills in addition to the knowledge elements) need to be studied to the same degree as tasks in the development of an acupuncture certification or license exam. They are to be analyzed, defined, and published; they need to appear on the job analysis questionnaire; and they need to be subjected to decision rules analogous to what is done with tasks. None of this appears to have been done.

There is an excellent reason for including knowledge and skill elements in the Certification Standards in addition to requiring the job analysis to enumerate the tasks that are done in the practice of the profession. Tasks may remain quite stable over the years but the accumulated knowledge of the profession might have practitioners engaged in vastly different kinds of practice over time in performing these tasks. In medical practice, for example, two tasks facing physicians might be to determine brain activity under given conditions or to treat a certain set of symptoms. To accomplish the first task, physicians would need to be aware of the newest brain imaging techniques so that they would ask for MRIs or CAT scans rather than EEGs; to accomplish the second task, they would need to know the newest drugs and their side and interaction effects so that they could take advantage of the latest pharmacological technology. Thus, identifying the tasks, while clearly necessary as part of a job analysis, is only part of the process. The same task may be appropriately accomplished in quite different ways over time since the knowledge base of the profession grows every year, and it is the knowledge elements, presumably the most current ones, that will appear on the test with the older knowledge elements regularly purged.

It is interesting that the directions for the job analysis questionnaire used by NCCAOM explicitly speak to the importance of knowledge and skill elements. The respondents are told, in part:

Those who are certified by the NCCAOM will be recognized to have the knowledge and skills necessary for safe and competent practice in one or more

of these modalities. The knowledge and skills expected of this practitioner include questioning, assessment, analysis and diagnosis, treatment, and evaluation of treatment...

One would therefore make the presumption that at some point in the examination development process knowledge elements were derived from the task statements and transformed into test questions. But because the knowledge elements were not included on the questionnaire, they were not rated by the sample of practitioners who completed the survey. Further, because they are not documented in any of the material that was provided by NCCAOM, there was no opportunity for this writer or anyone else outside of the organization to review and evaluate these knowledge elements (assuming that they were developed in the first place).

Evaluation of NCCAOM's Work

In my opinion, the failure of NCCAOM to include knowledge elements on the questionnaire, indeed, the failure to specify them at all in any part of their documentation, is in conflict with their own standards and is in conflict with what is ordinarily done in professional practice. This failure constitutes a serious breach in the chain of validity evidence, and may represent a serious flaw in NCCAOM's test development process.

Scales Used for Respondents to Evaluate Task and Knowledge Statements

The national job analysis presented a sample of NCCAOM acupuncture practitioners with the lists of tasks, whereas the occupational analysis done by OER presented a sample of California practitioners with content (performance) domains, tasks, and knowledge elements. The goal of both research efforts was to separate the important descriptions and underpinnings of the job that should serve as the basis of test development from those that were judged as less important and thus should not be part of the foundation of the licensing exam.

To accomplish this goal, respondents to the survey evaluate each item on the questionnaire on one or more response scales from which it is possible to assess the importance of the items. The particulars of the response scales used for this purpose have virtually infinite variation in their details, but they must all fall within the boundaries established by professional practice.

OER Occupational Analysis

The California occupational analysis asked respondents to rate the general content areas into which the tasks and knowledge elements were organized as well as the tasks and knowledge elements themselves. Content areas were evaluated on 4-point response scales of importance and criticality and on percent of time devoted to each area. Tasks were rated on 4-point response scales of importance, frequency, and criticality. Knowledge elements were rated on a 3-point scale of usage and a 4-point scale of criticality. All of the point scales were anchored at the low end by a value of zero representing *not important, never or very rarely, not critical*, and *I do not have the knowledge* for the importance, frequency, criticality, and usage scales, respectively.

NCCAOM Job Analysis

The NCCAOM job analysis contained only the tasks that practitioners performed. These tasks were rated on a 6-point response scale assessing the significance of the task anchored at 0 (*Not Performed*) and 5 (*Extremely Significant*). Respondents were instructed to consider both the importance of the task and the frequency with which the task was performed in judging how significant the activity contributed to effective practice.

Evaluation of the Two Procedures

The nature and quality of the rating scales by both research groups in assessing the importance of the respective items on their questionnaires are consistent with the standards established by professional practice. It should be noted that although California used more scales, these scales were statistically combined within the task section and within the knowledge section to yield a single value. The NCCAOM scale asked the respondents to make that assessment as an overall rating rather than asking for frequency and importance separately and later combining them statistically. Thus, the different approaches essentially reached roughly the same type of final judgment through somewhat different strategies. Both meet agreed upon professional standards.

Organizing the Tasks and Knowledge Elements

A lengthy list of tasks and knowledge elements can rapidly become unwieldy and disorienting. It is therefore common practice to superimpose on the list an organizational structure to categorize, at least on a global level, the specific tasks and knowledge elements contained in the list. Both OER and NCCAOM developed such an organizational scheme.

California's Task Structure

The process of determining the organizational structure of the tasks and knowledge elements was not addressed in the occupational analysis report. But in her March 10, 2004 letter to this writer, Dr. Ferrel indicated that the OER project manager sorted the task and knowledge statements and then had each of two 8-member SME panels review, critique, and modify this sorting as necessary. The five global categories resulting from this process were Patient Assessment, Developing a Diagnostic Impression, Providing Acupuncture Treatment, Prescribing Herbal Medicinals, and Regulations for Public Health and Safety.

NCCAOM's Task Structure

The process of how the tasks were organized was not fully addressed in NCCAOM's job analysis report. According to the report, the advisory SME panel thoroughly discussed the task list and, a reader would infer, discussed and eventually settled on categories to structure the tasks. At any rate, the final six categories that emerged were Questioning, Assessment, Analysis and Diagnosis, Treatment, Evaluation of Treatment, and Professional and Legal Issues.

Comparison of California's and NCCAOM's Task Structure

The global categories developed by OER and NCCAOM are quite similar. Evaluation of their comparability can be made at two levels: the global category labels and at the content area level or the individual task level. Note that we cannot compare them at the knowledge element level because the NCCAOM documentation does not address knowledge. Here, we will briefly discuss the global category level. In a later section we compare them at the content area and individual task level since some of the tasks in the set at this point will be removed because they were judged to be of insufficient importance to serve as the basis for examination development.

In comparing the global categories of tasks, it is useful to ask if we expect them a priori to be precisely the same. The answer is "not necessarily," and here is why. As is true for any large set of tasks, there are many possible organizing schemes that SMEs would judge to be valid representations of the profession. It would not be surprising, for example, if either agency gathered a fresh set of analysts and SMEs and attempted to organize the list of tasks from scratch that they would end with exactly the same categories that emerged the first time around. This is not to say that their previous and new categories would be terribly different; they would probably bear a very strong resemblance to each other. It is in this context that it is possible to say that the California and NCCAOM categories are, despite the apparent differences in the wording of their tasks, translatable into each other.

A more comprehensive approach is to evaluate the two occupational analyses at the content area and individual task level. This much more complex matching will be discussed later in the report.

Sampling Practitioner Respondents for the Job/Occupational Questionnaire

The purpose of administering a job or occupational analysis to a large number of current practitioners is to determine which tasks and knowledge elements in the comprehensive list are important enough to use as the basis of the licensing exam. The judgment how important is each task and knowledge element is made by practitioners; the test developers can then set a criterion of importance to be achieved for a task or knowledge element to be used in the creation of the examination.

Because the initial judgment of importance is made by practitioners, and because both test development groups polled a sample rather than asking every practitioner, it is very important to make sure that the sample is representative of the population of practitioners in the geographic arena in which the test is to be used.

The sample acts as a proxy for the larger group and must therefore comprise a miniature version of it. We assume that the sample will behave in the same way that the entire set of practitioners behaves. Our dilemma, of course, is that we do not know how the entire set of practitioners would behave if we could actually observe them and so we have no basis to guarantee that the sample is acting appropriately. Thus, when we ask practitioners in the sample to judge the importance of these various tasks and knowledge elements and compile their responses, we infer that these judgments are representative of all the practitioners in the population.

Since OER and NCCAOM cannot or do not access the entire population, they must attempt to configure the sample in such a way that it resembles the population. In terms of respondents to job questionnaires, the configuration is typically based on the demographic composition of the sample. There is a considerable amount of demographic information that can be obtained from practitioners, and test development groups will select the subset of variables they feel are important enough to measure in their particular job questionnaire.

Since we do not know how the demographic variables are distributed in the population, what we look for is diversity or dispersion on these demographics within the sample. For example, we would want the sample to contain practitioners representing a wide range of years since they were granted a license, we would want them to be practicing throughout the entire range of geographic locations covered by the license or certification, and so forth. We would also want the sample to contain a good mixture of men and women as well as the many different cultural backgrounds represented in the practitioner population. In short, we want the sample to look like a miniature version of the entire set of practitioners. To the extent that they are a small mirror of the population, the case can be argued that they would also behave in a manner similar to that population.

The key to making a strong case is selecting demographic variables that are both relevant to and reasonable to ask of those practitioners completing the occupational analysis. Since there are more demographic variables fitting these criteria than can readily be measured, each agency must make a decision regarding which ones it will attempt to capture.

California Sampling Process

While it is common for job or occupational analyses to be mailed to potential respondents, OER opted for a different strategy. “In an effort to increase the response rate obtained in this study, all practitioners who had been licensed twenty years or less were invited to come to one of 18 on-site panels held at various locations throughout California to complete the questionnaire” (Occupational Practice, State of California, Department of Consumer Affairs, November 2001, p. 3). A total of 290 practitioners attended and completed the questionnaire. Of these, 282 indicated that they were currently practicing the profession and so comprised the respondent sample.

The questionnaire queried respondents on the following ten demographic variables that were used for the occupational analysis: type of business (sole owner, partnership, corporation); hours worked each week; years licensed; work location (urban, rural); primary focus of practice (e.g., general, pain management); primary spoken language, languages spoken fluently, and primary language of patients (all with the choices of English, Chinese, Korean, Other); school at which acupuncture program was completed; and county of primary practice.

The results indicated, briefly, that most were the sole owner of their business, were in general practice, and worked in an urban location. Most spoke fluent English and the primary language of most of their patients was English as well. Respondents worked a full range of hours each week and represented a wide range of years that they were licensed. About a quarter of the respondents practiced in the general San Francisco Bay area and more than half were from the Los Angeles area. Roughly around 70% practiced in communities that are on or very near the coast of California.

Evaluation of the California Sample

According to the Sunset Review Report of 2001 (photocopied from a document in the Little Hoover Commission files), a total of 6,809 licenses were active in California during the fiscal year of 2000/01. This figure represents the maximum size of the population since some of those individuals might not have been currently practicing or may not have been residing in the state. Since 8 individuals who were sampled but were not currently practicing comprised about 2.84% of the sample, and if we subtract this same 2.84% from the size of the population (about 193 individuals), we can estimate that 6,809 minus 193 or roughly 6,616 practitioners were in practice in California. The sample of 282 thus represents 4.26% of the estimated practitioner population.

We can compare this percent with what one would ordinarily obtain in such studies. According to Dr. Ferrel on page 6 of her March 10, 2004 letter, “Response rates obtained from occupational analysis surveys typically range from 12%–60%.” OER’s under 5% response rate falls far short of expectations. At the same time, a sample size of 282 is not exceptionally small. Further, the distribution of the demographic variables discussed above suggests that OER probably obtained a relatively diverse group of respondents. Unfortunately, neither the gender nor the cultural background of the practitioner was asked, but we do know the county in which these practitioners practiced.

While we do not know how the universe of California practitioners align themselves on the demographic variables that OER measured, we do have relatively objective population counts of the state. If we make the working assumption that acupuncturists are located in California in approximate proportion to its general population, then we may be able to evaluate this feature of the sample. We therefore obtained the population information for 2002 that was available from the California Department of Finance's website with the understanding that the figures, in terms of proportions of the population residing in California counties, would be reasonably applicable to the prior year in which the OER occupational analysis was done.

Table 2 presents the results of that analysis. All counties with at least one respondent in the sample are listed together with their population, the percentage of the total state population represented by those counties, and the percentage of practitioners from those counties represented in the sample. The last column in the table shows the difference between the percent in the sample and the percent of total population residing in each county. We subtracted the total population percent from the sample percent so that a positive value indicates that the county was over-sampled and a negative value indicates that the county was under-sampled (given our above assumptions).

As can be seen from Table 2, while the representation in the sample is very close to that of the general population for the smaller counties, differences in some of the larger counties can be seen. It is noticeable, if not surprising, that two major urban counties, Los Angeles and San Francisco, are considerably over-sampled at 10.71% and 8.37%, and Orange county is not far behind at 5.43%; alternatively, it is also possible that proportionally more practitioners have established their businesses in these counties. Further, these may represent locations where the panels were scheduled or where public transportation provided most convenient access to the panels. Overall, though, despite somewhat of a bias toward these very prominent population centers and to the coastal counties in general (at least in central and southern California), there is clearly an effort to achieve geographic diversity that was largely met. Ultimately, since practitioners are not compelled to attend these panels, OER (as is true for all test developers) must rely on the cooperation and good will of practitioners to donate a half day or more of their valuable time to participate in such a project. In this respect, in my opinion OER did an excellent job of getting the representation that it did.

Table 2
The California Sample and State Statistics
For the Occupational Analysis

Geographic Region	Total Population^a	% Total	% Sample	Diff^b
Alameda	1,488,000	4.21	5.7	1.49
Contra Costa	989,800	2.80	2.5	-.30
El Dorado	165,400	.46	0.7	.24
Glen	26,950	.08	0.4	.32
Los Angeles	9,889,300	27.99	38.7	10.71
Marin	250,100	.71	1.8	.09
Mendocino	88,300	.25	.4	.15
Monterey	413,700	1.17	1.1	-.07
Napa	128,900	.36	.4	.76
Nevada	96,000	.27	.4	.13
Orange	2,959,300	8.37	13.8	5.43
Riverside	1,682,500	4.76	1.1	-3.66
Sacramento	1,302,600	3.69	1.4	-2.29
San Bernardino	1,816,500	5.14	.4	-4.74
San Diego	2,944,500	8.33	4.3	-4.03
San Francisco	788,200	2.23	10.6	8.37
San Joaquin	607,800	1.72	.7	-1.02
San Luis Obispo	255,400	.72	.4	-.32
San Mateo	711,700	2.01	1.1	-.91
Santa Barbara	408,400	1.16	1.1	-.06
Santa Clara	1,717,000	4.86	3.2	-1.66
Santa Cruz	259,100	.73	2.8	2.07
Solano	411,600	1.16	.7	-.46
Sonoma	470,700	1.33	.7	-.63
Tuolumne	56,500	.16	.4	.24
Ventura	788,200	2.23	.4	-1.83
Yolo	180,100	.51	.4	-.11
State of California	35,336,000			

^a Based on California Department of Finance, Demographic Research Unit Website, February 2004, E-2. *California County Population Estimates and Components of Change. Revised July, 1, 2002 and Provisional July 1, 2003, Table 1.*

^b Difference score is computed as % Sample – % Total.

Recommendations to California

It appears to me that OER did not succeed in achieving its goal of increasing the response rate in this study by using on-site panels, and it would be appropriate for them to recruit a larger sample in future occupational analyses. If OER wishes to continue using on-site panels, it is recommended that they also mail their survey to all of the acupuncture license holders in the state (and screen for duplicate questionnaires so that those completing an on-site survey do not also contribute a mailed survey as well). If the mailing turns out to be very expensive, then eliminating one of the 18 on-site panels and using those funds for a mass mailing should make such a mailing economically feasible. If possible, gender and ethnicity

or cultural background information should also be collected. Such information is an important component of the validity evidence supporting the occupational analysis and, therefore, the a component of the validity evidence supporting the entire testing process. Although it is recognized that OER is prohibited from collecting extraneous or irrelevant information on individuals serving as SMEs or as respondents to occupational surveys, gender and cultural background data are quite relevant and appropriate variables to measure in the present context.

NCCAOM Sampling Process

According to NCCAOM's job analysis report, there were 12,287 credentialed records in its data files. This represented 9,830 individuals because many of its credential holders held more than one credential. A questionnaire was mailed to a random sample of 3,500 credential holders, with the constraint that states with fewer individuals were selected. Responses were obtained from 1,265 individuals, about 12.87% of NCCAOM's national base.

The questionnaire asked for information on 16 demographic variables (plus four policy questions that are not relevant to the present work). The demographic variables on the survey were: how respondents classified themselves and what they used in their practice; the number of patient visits they experienced per week and the number of hours per week that they practiced; primary and other work settings; the state in which they practiced; years in practice; primary role, secondary roles, and source of training; NCCAOM and other credentials held; primary tradition of practice; and, optionally, their gender and ethnicity.

The following brief profile of the respondents summarizes the findings. Approximately 86% of the sample considered themselves to be an oriental medicine (50%) or an acupuncture (36%) practitioner primarily performing acupuncture and a fair amount of lifestyle counseling and Chinese herbal medicine activities. They generally had between 10 and 50 patient visits per week with most between 11 and 30, and tended to practice between 11 and 40 hours per week. About half were in a private office and about a third were in some kind of group practice setting.

As is common in nationally-based sampling plans, the entire country is divided into geographic regions. The two regions most heavily represented in the sample were the northeast (21%) and the southwest (18%). The other regions accounted for between 10% and 13% representation. On average, the respondents were in practice for almost 8 years but the distribution of the sample was strongly toward those who had practiced for 5 or fewer years. Virtually all of the respondents were practitioners who were formally trained in the United States. Of those who responded to the two optional questions, about two-thirds were male. In terms of ethnicity, two-thirds were Caucasian and a quarter were Asian, Asian American, or Pacific Islander.

Evaluation of the NCCAOM Sample With Respect to California

NCCAOM engaged in the more traditional procedure of mailing their job analysis questionnaire to their constituents. Their response rate was 36.14%, a very respectable outcome. However, we are not told of how well the sample matched whatever demographics are known about the entire population (e.g., how many certificate holders are in each state or region). The random sample selected for the mailing, because of its absolute size and the fact that it covered a third of the population, very likely was quite representative. But NCCAOM received back only a third of the mailings and thus heard from just under 13% of the population; this group could very well have been less representative. For example, there were relatively few individuals who had been in practice for a long time. Of course, the sample participating in the California occupational analysis conducted by OER is open to the same concern.

All told, given the sheer size of the sample, the fact that so many different states and regions were captured, and the wide range of responses obtained on most of the demographic questions including the optional gender and ethnicity items, my assessment is that for a nationally-based exam, the sampling was probably sufficient to capture a representative cross-section of the national population.

But the focus of this report is somewhat different than NCCAOM's emphasis. The central issue here is the suitability of the national exam process for use in California. The information received from NCCAOM on April 30, 2004 contained the breakdown of those individuals in the national survey who were from California. According to that information, there were 179 practitioners based in California. Using the estimate of 6616 licensed practitioners in the state, the national sample appeared to contain 2.71% of the estimated California practitioner population.

The demographic information collected by OER for the California survey and by NCCAOM for the national survey is quite different. For the California practitioners in the national survey, 22.3% considered themselves to be acupuncturists and 62.2% considered themselves to be oriental medicine professionals, for a total of 84.5%. This is very close to that found for the entire national survey. A total of 77.5% of the respondents indicated that the primary tradition of their practice was Traditional Chinese Medicine. There are no comparable questions on the OER survey.

Both groups were asked the number of hours they worked per week, and these matched up quite well as shown in Table 3. The largest difference is in the 21 to 30 hours per week category, and that discrepancy is not especially troublesome.

Table 3
Number of Hours Worked Per Week
For California Practitioners in the
NCCAOM and OER Surveys

Hrs Worked per Week	Percent CA Practitioners in the NCCAOM Survey (n = 179)	Percent CA Practitioners in the OER Survey (n = 282)
Under 10	11.2	11.0
11 to 20	18.1	16.0
21 to 30	27.4	17.7
31 to 40	28.2	30.7
Over 40	14.7	24.1

The national survey asked respondents how they obtained their training. Of the California participants, 90.6% received formal education in the United States and another 6.4% did an American apprenticeship. OER’s survey identified the school at which their respondents’ acupuncture program was completed. Of the schools listed (there is a category of “Other”), only one (represented by 1.4% of the sample) is located out of the country. It would seem, therefore, that the vast majority of those in the OER survey also received their education in schools located in America. Without further interfacing between the two surveys, however, it is difficult to make a more definitive comparison of the two California samples.

Another issue of interest here is the degree to which California is represented proportionally in the national sample. This is an appropriate consideration here since the job analysis defines the important portions of professional practice which in turn drives what appears on the test. To begin answering this question, we can simply note the size of the two populations and see if we can estimate their overlap.

There are 9,830 individuals in NCCAOM’s data base and 6,809 license holders in California. How much overlap there is between these two populations might have been told to us if we would have been able to perform the double test taker analysis, but we still may be able to make a very rough guess at this figure. The 179 California practitioners captured by the national survey comprise about 14.15% of NCCAOM’s sample of 1,265. If we assume this percentage is close to California’s representation in the entire data base, then of the 9,830 certificate holders about 1391 of them (14.15%) would be California practitioners. Subtracting them from the 6,809 California license holders leaves 6,809 minus 1391 or 5,418 California practitioners presumably not contained in the NCCAOM data base.

If California adopted the national exam as its licensing test battery, then California’s representation in the national job analysis should be considerable. NCCAOM may have about 1391 of California’s practitioners in their data base right now but there would be another 5,418 that would be new to their job analysis domain if the national test were attempted to be applied to California. That is, the practitioner base for which NCCAOM would need to account in their job analysis would jump over 50% and be dominated by the huge percentage of California practitioners. In approximate (and probably conservative) numbers, the number of practitioners comprising the population from which a job analysis

questionnaire must then sample would increase from the 9,830 existing individuals to approximately 15,248. About 6,809 of them, or almost 45%, would be based in California.

The point of this numerical exercise is as follows. Under the hypothetical scenario in which California adopted the national testing process, the sampling plan for a national job analysis would have to be very different from the one that was implemented for the existent job analysis. In order for California to place the necessary confidence in a NCCAOM licensing examination process, the job analysis would have needed to ensure that the way in which acupuncture and oriental medicine is practiced in this state was proportionally represented in the job analysis which, it must be remembered, drives the content that appears on the test. At this moment in time, the 179 California practitioners represent only 14.15% of the job analysis sample, a proportion that does not appear to be sufficient to provide the necessary level of confidence needed for the Acupuncture Board to immediately use the NCCAOM test for licensing. If they or the California Legislature opt for using the national test, they should do so only after a new job analysis with appropriate California representation is completed and after any revisions based on that job analysis were made to the testing modules.

Distinguishing the Important Tasks and Knowledge Elements

Once the responses to the questionnaire items are statistically analyzed, it is necessary to separate the important tasks and knowledge elements from the universe of tasks and knowledge elements that comprise professional practice. The latter represent the full description of practice including those portions that are not especially important, while the former represent the essence of the profession. It is this former set, the important portions of the job, that should serve as the basis of examination development. Both research teams, OER and NCCAOM, established procedures to identify the important components of the profession.

California's Strategy to Identify the Important Job Components

OER developed what it called "criticality indices" to gauge the importance of the job components. It then established a threshold value that each component needed to obtain in order to serve as the basis of the license exam. Although the occupational analysis report is missing much of the documentation of this process, the letter from Dr. Ferrel dated March 10, 2004 clarified a good deal of what was missing. My report is based on information obtained from both sources.

A task criticality index was computed by OER as follows. For each rater, his or her importance, frequency, and criticality ratings on each task were multiplied together. These were then aggregated across raters. A panel of SMEs then reviewed the index for each task and, I assume, together with an analyst from OER decided on a threshold value for the "survival" of a task. A knowledge criticality index was computed in an analogous manner.

For the knowledge criticality computation, for each rater the usage and criticality ratings were multiplied together and the panel decided on a threshold for these elements.

This process resulted in 150 tasks and 251 knowledge elements being brought into the test development phase.

Evaluation of California’s Strategy to Identify the Important Job Components

It is fairly common practice to combine individual ratings into some sort of a composite criticality index, and multiplying rating values is often done. It is, in my experience, more common to multiply the ratings once they have been averaged across all respondents, partly because it “even out” extreme scores from a very few respondent and partly because it avoids a technical glitch to which OER fell prey.

The scales respondents had to use to judge the tasks and knowledge elements contained the value of zero. Zero means “absence” of something. It is an absolute judgment and should be used only if absence makes sense for what is being rated. One could argue that a Frequency rating, asking for how frequently a task is performed, can take on a zero value. That would mean that the task is never performed. But most of the other scales really do not fit well with a zero value. For example, even if a task is felt to be unimportant, we are talking about a relative rather than absolute judgment. Thus, the lowest value on the scale is usually given as a 1 rather than a zero.

The issue of using zeros or 1s as the lowest scale value is often not heavily debated because the scales are averaged over respondents prior to the multiplication operation. Thus, even if a respondent assigned a zero to the importance of a particular task or knowledge element, the importance ratings of all respondents would be combined to acquire an average importance rating for each component. That average would be multiplied by the average of the other ratings to compute a criticality score for each task or knowledge element.

The technical glitch in the present instance occurred because OER multiplied the ratings of individual raters. In this way, a zero on any scale would result in a criticality score of zero for that task or knowledge element for that rater. This probably affected the outcome but in some way that cannot be specified here. This is true because OER presented their data in a form different from the way they calculated criticality. That is, they showed in their tables the average of each scale across raters; if one does the multiplication, one arrives at a value different from the criticality score that is actually reported for each item. While this does not comprise a fatal error in the data processing, it is not good testing practice and it is antithetical to the need to document what was done.

NCCAOM's Strategy to Identify the Important Job Components

NCCAOM focused only on tasks in their job analysis, thus missing part of what they needed to evaluate according to their own standards. Essentially, they applied in a sequential manner a set of decision rules that filtered the task list multiple times to achieve their importance threshold.

Without describing in detail each of these decision rules, consider one example. The first pass through the data asked if the task was performed by a majority of the respondents. In an illustration of the proper way to use zero ratings, 59 tasks were eliminated because they were not performed (were given a rating of zero) by more than 50% of the respondents. Other decisions were based on significance to practice in general and regardless of size of practice, years of experience, and geographic region. This process filtered out 72 of the tasks to bring the total to 212 important tasks taken into the test development phase.

Evaluation of NCCAOM's Strategy to Identify the Important Job Components

The strategy NCCAOM used to identify the important tasks was indicative of a high quality process. Successive use of decision rules resulted in a set of tasks that were clearly important to the practice of acupuncture and oriental medicine. But the test items were constructed directly from the task statements with no knowledge elements identified as an intermediate step.

Reliability of the Task and Knowledge Elements Ratings

The Concept of Reliability Concerning Job Or Occupational Analysis Surveys

Reliability is one of the quality assessments that we use to evaluate a test, and we discuss this concept later in the report. In the context of an occupational analysis, reliability addresses the consistency of the responses we obtain on a job survey. Very low reliability means that there is little consistency; in its most exaggerated form, it is akin to random responding. Very high reliability means that the responses are very consistent. From an evaluation standpoint, it is necessary to see high values of this index.

There are two kinds of consistency that can be assessed in analyzing occupational analysis data: consistency in the responses across items and consistency in responses across respondents. Item consistency is commonly indexed by a statistic known as Cronbach's coefficient alpha and assesses what is called internal consistency (of the responses to the survey items). Respondent consistency is usually indexed by a statistic known as the intraclass correlation and assesses what is called rater reliability.

Internal consistency essentially is a function of the correlations among the items. If the items on the job analysis represent different kinds of content, which they very well might,

they may not be all that highly associated with each other. This would tend to lower the value of coefficient alpha. If the items represent similar content, alpha might be higher. In either case, all that the calculation of coefficient alpha does is describe the nature of the content base, and either result may be perfectly acceptable. Thus, in most situations there is not all that much to be gained from the results of that analysis.

Of the two reliability indexes, by far the more important to assess for the job analysis data is rater reliability. Through this analysis we can evaluate the degree to which the respondents tended to see the content of their professional practice in similar ways. High reliability here tells us that the raters share a common viewpoint—where one practitioner believes something to be relatively unimportant, other practitioners judge likewise; where one respondent believes something else to be relatively important, other respondents see it in the same way. Lower respondent reliability tells us that there are differences of viewpoints among the practitioners; although that may make test development more difficult, it is an important fact to know. Extremely low respondent reliability can signal many possibilities including chaos in the profession, a poorly constructed job survey, poorly designed SME panels; in any case, it is a warning to stop the process and determine why the results are emerging as they are.

Reliability Results For the California Occupational Analysis

Reliability indexes were supplied by OER in the occupational analysis report but which type of reliability was computed was not at all specified. Remaining silent on this issue usually indicates that coefficient alpha was used but I did query OER on this matter. In her March 10, 2004 response to my inquiry, Dr. Ferrel confirmed that the reliability index was Cronbach's alpha. Obviously, rater reliability was not computed. Coefficient alpha ordinarily ranges between 0 which shows no internal consistency and 1 which showing perfect internal consistency (although it is possible for alpha to take on negative values under some anomalous or unusual circumstances). Obtained reliability coefficients for tasks and knowledge elements ranged from .75 to .95 and represent relatively high alpha coefficient values.

Reliability Results For NCCAOM's Occupational Analysis

Unlike the California results, NCCAOM reported both types of reliability but only with respect to tasks since no knowledge elements were contained on the survey. Its job analysis talks about reliability for tasks and reliability for raters. The former is very likely computed as coefficients alpha; the latter is probably one of the variants of the intraclass correlation. However, in neither case is the name of the statistic specified. Task reliability ranged from .88 to .98 and rater reliability hovered around .99.

Evaluation of the Reliability Results

The relevant index here is rater reliability, which is usually quite high in well done job or occupational analyses. NCCAOM's results are about as high as the index can get but was done only for tasks since there were no knowledge elements contained on the survey. California's occupational analysis did not assess rater reliability; it is recommended that they do so in future occupational analyses.

Linking Tasks and Knowledge Elements

It is appropriate to link the tasks and knowledge elements to ensure that there are no free floating components when building the licensing exam. That is, tasks with no associated knowledge elements and knowledge elements that cannot be related back to a task should not be used in the test development process.

Since only the California occupational analysis identified knowledge elements, this was the only test development team that could accomplish this linking process. OER used a panel of SMEs to provide their professional expertise in this endeavor, and documented their work thoroughly in a series of tables provided in their report. Their work is a fine example of best practice.

Stage 2: Examination Development

Weighting the Content Areas

The Weighting Process Used By California and NCCAOM

In order to judge that candidates meet the criterion of minimal competency in acupuncture and oriental medicine, the license examination must be a representation of the various aspects of the profession. By analogy, a written test for the driver's license must cover more than braking distances and sign interpretation to be a valid measure of the requisite knowledge we expect of motor vehicle operators. Thus, it is necessary that the full range of the competencies underlying acupuncture and oriental medicine be tested in proportion to the role they play in practice. This proportionality requirement cannot be overstated in terms of importance.

The proportionality standard is captured in the development of a licensing examination by the weighting structure of its components. There are several different strategies that can be used to reach a valid weighting structure. OER based the weighting of the CALE on the representation of tasks and knowledge elements within each of the major categories shown in the occupational analysis. The CALE is structured to test five content areas as shown in Table 4. It consists of 200 test questions. A small number of these are pilot items written to be used in future exams but that are not scored. The percentages of items in each content are

shown in the table. These percentages are also meant to represent the “apparent” weights of the different content areas.

Table 4
Weighting Schema for the
CALE’s Content Areas

Content Area	% of Total
Patient Assessment	25%
Developing a Diagnostic Impression	20%
Providing Acupuncture Treatment	29%
Prescribing Herbal Medicinals	17%
Regulations for Public Health and Safety	9%

NCCAOM derived the weights for their modules through a complex but very effective iterative process that involved their advisory committee, the NCCAOM Board and Examination Committees, and focus groups that represented some state license agencies.

The NCCAOM test structure for acupuncture and oriental medicine is divided into 5 modules as shown in Table 5. These modules contain different numbers of items that essentially define the overall “apparent” weighting schema, and these are also shown in the table. Should the NCCAOM exams be adopted for use in California, the modules need to be combined in a weighted fashion (unless a modular “required pass” strategy is adopted). Based on the current job analysis, which would need to be revisited if California was absorbed within the national program, the weights for each module are shown in the last column of Table 5. I calculated these from the number of items used in the combined set of modules which is 410. For example, the acupuncture module contains 90 questions and is about 22% of 410.

Table 5
Weighting Schema for
NCCAOM’s Five Modules

Test Module	# of Items	% of Total
Foundations of Oriental Medicine	125	30.49%
Acupuncture	90	21.95%
Point Location	25	6.10%
Chinese Herbology	120	29.27%
Biomedicine	50	12.19%
Set of All 5 Modules	410	100.00%

Each test development group, in its own way, derived a valid and defensible “apparent” weighting schema for its exam.

The Missing Element in These Weighting Schemas

The weighting schema for both California and the national exams was referred to as an “apparent” weighting schema. We are using the notion of “apparent” to indicate that these are the weights built into the structure of the tests by virtue of the number of items used to represent each area. But the actual weights of these areas in terms of their contribution to the total test score could differ from these structural weights. This is because the actual weights depend on the variance of the responses on these content areas, and neither group addresses this in their test analysis.

To see why variance can impact a weighting schema, consider a simplified and extreme example. Assume that an exam tests two topic area, Topic 1 and Topic 2. It has been determined that Topic 1 should be weighted 35% and that Topic 2 should be weighted 65%. A 100 item test is built containing 35 questions to test the first topic and 65 questions to test the second topic. At this point, the 35% – 65% breakdown is the “apparent” weighting schema, the one the test developers consider to be valid and therefore the one that they want.

Now they administer the test and score the items. In our example, we will look at the scores for each topic separately for three candidates. Candidate A answered all 35 Topic A questions correctly but answered only 30 of the Topic B questions right. Candidate B answered 20 Topic A questions and 30 Topic B questions correctly. Candidate C was correct on 10 Topic A questions and 30 Topic B questions. When we obtain a total score for each candidate, we find Candidates A, B, and C received scores of 65, 50, and 40, respectively.

If we examine these results in terms of weights, we find something disconcerting. The scores of the candidates—the weight of each topic—is all about their performance on Topic 1, the area that was presumably weighted only 35% of the total. Topic 2, which was “apparently” weighted 65%, contributed nothing to distinguishing among the candidates. If you were predicting the total score Topic B should have been weighted 65% if it were to meet the requirements of the test developers but in reality it has an effective (actual) weight of zero.

The reason that this occurred in our simplified and extreme example is that there was no variance in the scores for Topic 2. Variance contributes information and only variables that contain information can be meaningfully weighted.

While this example is unrealistic, the point that it shows is quite real. The weighting schema that you start with—your apparent weights based on the number of items in each content area—is not necessarily the functional weight they play in computing the total test score. Thus, one of the pieces of information that you need to examine, and a piece of information which should be documented in the test analysis, is the variability of the candidate responses associated with each content area once the test questions are scored. To the extent that the variability is comparable, the content areas would not gain or lose much in the way of their apparent weights based on the number of questions representing them in the test. To the extent that the variability of the scores on some content areas are constrained, the final test score could represent a different effective weighting schema despite the best intentions of the test developers. Both test groups would do well to incorporate this consideration into their test analysis processes.

Content of the Examinations

The following is a somewhat more elaborated description of the structure of the tests.

Structure of the CALE

The CALE consists of 200 test questions covering five general content areas listed below. Of these 200 items, 175 are scored for the examination and 25 are pilot items to be evaluated for use on future exams.

- Patient Assessment (25%) covers such topics as obtaining a patient history, performing a physical examination, evaluation for western pharmacology, and diagnostic testing.
- Developing a Diagnostic Impression (20%) covers diagnosis, differentiation of syndromes, biomedical disease, and treatment planning.
- Providing Acupuncture Treatment (29%) includes point selection principles, categories, location, and needling techniques. It also covers performing auxiliary treatment including Microsystems, treatment observation and modification, and treatment contraindications.
- Prescribing Herbal Medicinals (17%) covers identifying herbs as well as prescribing and administering herbs.
- Regulations for Public Health and Safety (9%) tests the laws and regulations governing hygiene and controlling pathogenic contaminants, maintaining patient records, and reporting suspected child, elder, or dependent adult abuse.

Structure of the NCCAOM Tests

The new NCCAOM examination structure, launched in early 2004, offers a set of five written test modules that, taken in combination, appear to cover the range of content tested by the CALE. It would thus appear that if the NCCAOM examinations were to be used for California licensure purposes, candidates would likely take all five modules unless the structure of the license was substantially changed from what it is now. These five modules together comprise 410 test items. Percentages provided below are based on this full set of 410 questions.

- Foundations of Oriental Medicine (125 items, approximately 30.5%) covers patient history, assessment of patient status, analysis and diagnosis, specifying a treatment plan, patient record keeping, ensuring equipment maintenance and safety, and implementing infection control precautions.

- Acupuncture (90 items, approximately 22%) tests diagnostic treatment and planning, points including sets of points and microsystem points, treatment considerations (including moxibustion and heat/cold), clean needle technique, and evaluating treatment results.
- Point Location (25 items, approximately 6%) deals with identifying the acupuncture point corresponding to indicated locations on a diagram of a human body.
- Biomedicine (50 items, approximately 12.2%) covers patient history and status, consideration of the results of western medical tests and assessment, reporting requirements for such issues as abuse and disease transmission, maintaining professional hygiene, and a variety of safety issues.
- Chinese Herbology (120 items, approximately 29.3%) tests identification, selection, and formulation. It also covers applying herbal treatments, identifying precautions and contraindications and various interactions, and evaluating the results of treatment.

The topics tested by both of these examinations are very similar. For example, obtaining a patient history is covered in the Patient Assessment area of the CALE but is also tapped into in the NCCAOM exams in Foundations of Oriental Medicine and Biomedicine modules. Diagnosis is found in Developing a Diagnostic Impression but is tested in the NCCAOM exams in Foundations of Oriental Medicine and Acupuncture modules. Knowledge of herbs is tested by Prescribing Herbal Medicinals in the CALE and by the Chinese Herbology module from NCCAOM. Safety issues are addressed in the Regulations for Public Health and Safety portion of the CALE and by the Foundations of Oriental Medicine, Acupuncture, and Biomedicine modules from NCCAOM. Knowledge of points is found in the Providing Acupuncture Treatment area of the CALE and in the Acupuncture and Point Location modules of the NCCAOM test set.

Mapping the California and National Content to Each Other

Overview of the Mapping Process

One of the more relevant issues that arises in a comparison of two tests that are considered to be used as a licensing examination for California is whether they encompass the same content. This issue can be addressed at two levels: more general and more specific. At the more general level, one can ask if practitioners in California engage in similar tasks to those in the other states represented in NCCAOM's job analysis. Such an analysis addresses the way in which practitioners define their professional domain. If it is determined that the two sets of practitioners differ at this level, it would be difficult to argue that the tests are fully interchangeable and, while it could prove interesting, it is less clear what the analysis at the more specific level could tell us.

The Effect of Having No Knowledge Elements From the National Job Analysis

If the analysis at this more general level shows a comparability of content, the tests will have partially proven themselves capable of being used for the same purpose. But achieving comparability here does not mean that the tests can be treated as interchangeable. As was discussed earlier, tasks can remain intact over relatively long periods of time but the knowledge underlying those tasks may reveal professions that are quite different. This is true for at least two reasons. First, the knowledge base may change quite a bit over time. A license examination testing the state of knowledge at Time 1 may be outdated, and thus not valid, at Time 2. Second, it is conceivable that the profession may focus on the same tasks but use very different knowledge bases at the national level and at the California level to accomplish them. A license exam focusing on the knowledge base for one set of knowledge elements may not be appropriate to use as a license exam for the other set of knowledge elements.

This second reason may have some applicability to the present situation. Consider an extremely simplified hypothetical example. Assume that practitioners of a certain profession engage in the following two tasks at both the state and national levels: take a personal history and identify personality characteristics. On the surface these two geographic regions, which focus on what appears to be the same content, might be thought of as comparable and a licensing exam used for one might very well be proposed as being used for the other.

But despite the fact that the tasks appear to be the same, the knowledge base of these two disciplines might be found to be very different. For example, it might be found that the California practitioners take a personal history to understand the developmental dynamics that shape the way children will learn to deal with their world. This personal history will be driven by that knowledge base and inquiries will be made about childhood experiences and feelings. Assume that practitioners outside of California, on the other hand, might take a personal history to determine the alignment of the stars and planets under which the person was born and so may inquire about the exact time and location of the client's birth.

It is not claimed that the practice of acupuncture and oriental medicine finds itself in such a dire or extreme state as described above. Rather, it is very likely that the way in which acupuncture is practiced in California is relatively similar to the way in which it is practiced in other states. But the point needs to be made forcefully that even if the two populations of practitioners engage in what appears to be the same tasks, there is no assurance on two issues that the same licensure test can be used for both. First, until the analysis is done, we would not know if the knowledge bases underlying these tasks completely overlap. Second, if they do not overlap completely, we do not know if the state and national examinations test only the overlapping content or if they test knowledge unique to one population but not relevant to the other.

This is why the second level of analysis done at the more specific level—the level of the knowledge elements—is so important to perform. Unfortunately, NCCAOM's job analysis did not document the knowledge elements underlying the tasks they identified. We therefore cannot perform this analysis. All that can be done is the analysis at the more general level of tasks, and that is what is reported in this section of the report.

The Structure of the Two Content Domains

California's Structure

California's occupational analysis organized the tasks that were identified into a three-level hierarchy. At the most global level are the five content areas of Patient Assessment, Developing a Diagnostic Impression, Providing Acupuncture Treatment, Prescribing Herbal Medicinals, and Regulations for Public Health and Safety. Each of these main content areas were divided into subareas. The smallest number of subareas totaled two for Prescribing Herbal Medicinals and Regulations for Public Health and Safety, and the largest number of subareas was seven for Providing Acupuncture Treatment. Within each of these subareas were the tasks generated from the occupational analysis. In all, there were 150 tasks at this level of the hierarchical structure reaching the state's criticality criterion for inclusion in the test specifications.

NCCAOM's Structure

The tasks identified in the national job analysis were organized into a hierarchy of four different levels. The highest order of this hierarchy was the module level, and the five modules were Foundations of Oriental Medicine, Acupuncture, Chinese Herbology, Biomedicine, and Point Location.

With the exception of the Point Location module, where it was stated that all of the points may be assessed but was otherwise not further broken down, the four other modules were divided into content categories. This structure comprises the second level of the organizational hierarchy. The number of content categories varied somewhat by module, ranging from three categories for the Biomedicine and Acupuncture modules to five categories for the Foundations of Oriental Medicine module.

The third level of NCCAOM's organizational hierarchy represented content subcategories. These described more specific aspects of the content categories and in many instances appeared to be, in some sense, a label for a general task that practitioners would perform. For example, in the Foundations of Oriental Medicine module there is a content category titled "Assessment." One of the subcategories under Assessment is "Looking."

The fourth and most specific level of this organizational hierarchy represented what was called in the job analysis the "tasks" of the profession. For example, under the subcategory of "Looking," tasks such as the following appeared: "eyes," "body structure," "spirit/shen," and "symptom site/local area of complaint." The majority of the elements at this fourth level of the hierarchy, although certainly not all, were analogous to these in that they were in an abbreviated form as these and contained virtually no verbs in their expression. Overall, there were 212 tasks represented at this level of the hierarchy that passed the criteria established by NCCAOM to serve as the basis for the test specifications.

Although the elements listed at this most specific level are identified as tasks in the national job analysis, it is an uncommon use of such a characterization. Tasks are usually

thought of as actions or behaviors with some explicit purpose in mind. As a contrast, OER's occupational analysis also identifies the eyes of patients as worthy of assessing. But the wording of California's task statement is as follows: "Observe the face and eyes to determine the outward manifestation of the Shen." This latter statement is a more traditional (and, in my view, much more appropriate) way to express a task in that we understand what behavior is being described and the reason for engaging in that behavior.

The Process Used to Map the Two Content Domains

Direction of Mapping

Although at the most specific level of each organizational hierarchy are elements called tasks by both test development agencies, these tasks differ substantially in the way that they are expressed. The wording in the national job analysis is, by and large, much more abbreviated than what we see in the state's document. This lack of specificity at the national level therefore introduces a certain amount of uncertainty into any matching procedure.

The strategy used here was influenced by the fact that (a) the national structure was "deeper" than that of the state, and (b) the state's task statements were generally more complete than that of the national's statements. It therefore turned out to be more feasible for us to use the national structure as a base and to map the state tasks to it rather than the reverse. At the end of the day, though, we expect that the strategy we used should not make that much of a difference; that is, we believe that we would have obtained a very similar result if we had mapped NCCAOM's tasks to the structure used by California.

Some Limitations of Our Process

We implemented this mapping strategy by attempting to identify tasks that appear, at least on the surface, to address the same sort of content. Without a comparable level of specificity in the task expressions, and certainly without the presence of knowledge elements from the national job analysis to guide us in this matching procedure, it is recognized that our surface matching process might have missed the mark on several occasions. It is possible that some of these misses might have been avoided had we made use of a practitioner panel to accomplish this mapping. However, we made an early and explicit decision to not involve practitioners in the preparation of this report. Our decision was based on the concern that we might give the appearance that our recommendations were potentially influenced by practitioners who might be thought to have already formulated opinions about the two testing programs that were being evaluated. Thus, we chose to permit some degree of uncertainty or error into our mapping process but at least take full ownership of it rather than give even the appearance of practitioner influence on the technical content of this report.

The Results of the Mapping

The goal of this analysis was to map the 150 tasks statements identified by the state to the 212 task elements identified by NCCAOM in their respective occupational analyses. This was accomplished by isolating the subcategories of the national list and considering each of the task elements within each of those subcategories. We then searched the California task list for those appearing to us to correspond with NCCAOM's subcategory and/or task elements.

We present the results of this rather large analysis in a series of tables which are contained in the Report Supplement at the end of this report following the attachments. Each table represents a different NCCAOM subcategory, and this subcategory and its associated task elements are listed in the left column. Those corresponding tasks from the California occupational analysis are presented in the corresponding row of the right column of the tables. Where possible, we matched the California task statement to the national task element. When we could not match specific elements but judged that the California tasks should be linked to NCCAOM's subcategory, we placed them at the bottom of the table.

There were three subcategories in the national test specifications for which we could not match any state tasks. These three subcategories were Apply Heat/Cold, Apply External Herbal Applications, and Treat Patient Using Nutritional Supplements and there appeared to us to be no matches from the state list.

Two subareas of the state's hierarchy of tasks did not appear to match the specific tasks elements identified by the national job analysis. One of these was "Forming a Diagnostic Impression." This subarea included tasks such as, "Integrate symptoms of physiological systems to determine the stage of disease progression," and "Differentiate between root and branch of disease by evaluating symptoms to determine focus of treatment." This set of tasks was matched to the content subcategories "Formulate Treatment Principles" and "Select Treatment Strategies." These subcategories are within the content category "Oriental Medicine Treatment, Planning, Principles, and Strategies" of the Foundations of Oriental Medicine module. We could not make a more specific match because the national subcategory contained no task elements (although the subcategory itself was matched to another California task from a different set). The tasks in the state's Forming a Diagnostic Impression set were nonetheless felt to be related to this same subcategory and so were placed at the end of table.

The other subarea of the state's hierarchy of tasks that did not appear to match any of the specific task elements identified by the national job analysis was "Point Selection Principles." This subarea included such tasks as, "Select local points by evaluating clinical indication to treat condition," and "Select points on the extremities to treat conditions occurring in the center." This set of tasks was placed on the table representing NCCAOM's content subcategory "Points and Sets of Points." This subcategory is in the content category "Diagnostic Techniques and Treatment and Planning" of the Acupuncture module.

The match here was made with much less confidence than any of the other mappings. The state tasks primarily deal with (a) using anatomical landmarks to locate points and (b) selecting points to treat various ailments. Outside of the Point Location Module which does not appear in the test specifications for the national exam, this subarea seemed to best fit within NCCAOM's Acupuncture module. Generally, four of the subcategories within this

module addressed task elements related to point principles and were therefore possible candidates for a match. One of these was already mapped to a state task. A second concerned microsystem points and therefore seemed less related. A third subcategory contained treatment theories and also appeared to be less related to the more general group of Point Selection Principles from the state's hierarchy. Finally, the fourth NCCAOM subcategory was Points and Sets of Points, and this is the one to which we ultimately made the match.

Evaluation of the Mapping Results

Although there were a few places within this mapping analysis, as just documented, where we have some uncertainty about our success, in the vast majority of instances we are confident that the task elements, or at least the subcategories, identified in the national job analysis correspond to the tasks that were identified by California. Thus, at the task level, we are relatively confident that practitioners in California seem to accomplish the same sort of tasks as those in the rest of country represented in NCCAOM's sampling.

Crafting Test Items

The processes thus far reviewed in this report have focused on how the test development teams for California and national exams have determined the subject matter of what should be tested in terms of both the content of the items and the proportional representation of that content and whether that content, at least at the level of the tasks, is comparable. Ultimately, that content needs to be translated into test items. In both cases, these items are pencil and paper multiple choice questions.

How well the process of translating the appropriate content into test items has worked can be approached from two directions. From the development direction, we can examine the process by which this is accomplished. A solid, well conceived process is likely to result in very good test items time after time. This will be discussed in the present section of the report. From the outcome direction, we can examine the statistical performance of the test items. High quality items will reveal themselves by yielding very favorable item and test statistics. We will discuss this outcome approach in the Quality Assessment section of the report.

It should be noted that the item creation process was documented in neither the test development report from OER nor that from NCCAOM. It was therefore necessary to contact both agencies to learn about their process. Both responded in March 2004 and the information they supplied served as the basis for these next two subsections of the report.

The California Process

In her response of March 10, Dr. Ferrel of OER outlined the process they use to produce test items. The total process takes six to eight months for each exam cycle. At the start of the process, item writing workshops are conducted with somewhat less than a dozen SME representing a balance of geographic regions and years licensed but with the provision that

“at least half of the practitioners who participate in the item writing workshop have been licensed five years or less to ensure examination questions are geared toward the entry-level practitioner.” SMEs are given extensive item writing instructions and then, under the supervision of a test development specialist from OER, prepare and revise the items. Two to three of these item writing workshops are conducted for each test followed by the same number of item review workshops. These latter workshops consist of about half a dozen SMEs who review and suggest modifications to the items. Only when items are passed without change are they entered into the item bank.

All items that are deposited into the item bank are pretested under examination conditions before they are used to contribute to a test score. Each license exam thus contains a small percentage of pilot items that test takers cannot distinguish from the real items. This procedure is exactly what should be done so that newly written items can prove their worth before they join the community of potential test items.

As of a letter dated April 1, 2004 from Dr. Ferrel of OER, the CALE item bank contains 1,777 items with up to about 150 items added to the bank each year. The number of new items added to the bank, according to Dr. Ferrel, “has been reduced due to fewer item-writing workshops as a result of reduced staffing levels at [OER].”

The NCCAOM Process

In its response to my inquiry, NCCAOM did not supply a description of its examination development process. Instead, its April 29, 2004 response to the question of how the task statements were translated into the format of a written test produced this two-sentence response:

As indicated by the content outlines and the job analysis report previously submitted for your review, tasks appearing on the job analysis survey, which met specific criteria, were retained and placed on the appropriate content outline. These content outlines serve as the basis for writing entry-level items.

NCCAOM uses a much less conservative strategy in placing new items in their exams. Whereas California will not score their pilot test questions the first time around, NCCAOM will score the item as part of the test unless they flag it for poor performance. This strategy is apparently a “screen out” rather than “earn its way in” standard. In their April 29 response, they stated the following:

New items are placed on the exam in Active Status. If an item performs unsatisfactorily, the item is not scored. The appropriate examination development committee then reviews the item before it may appear on another examination.

It was not made clear what the criteria were to define unsatisfactory performance but it probably means that the item’s difficulty level was .99 or 1.00 (see Attachment F, which is discussed more fully later in this report). It is also not clear whether the appropriate examination development committee will also rewrite the item or if it will just place it on another test if they feel the situation warrants such.

Based on the material I received on April 30, 2004, NCCAOM's item bank contains 2,285 items with about 1,000 items added each year.

Evaluation of the Two Processes

The process used by OER to craft new test items is, in my opinion, an example of best practice. It would be useful to have more new items written each year, but the current levels should prove adequate to maintain the integrity of the exam for the next few years. The process used by NCCAOM still remains undocumented.

In terms of evaluating new items, the California strategy fully protects candidates from being evaluated on pilot items. NCCAOM's strategy, while not inherently a poor process, depends upon their definition of "unsatisfactory" and the degree to which that definition fluctuates across modules or administrations of the same module. Without knowing this information, their procedure cannot at this time be endorsed. On the other hand, NCCAOM does turn its items over at a somewhat more rapid rate than California, which probably increases the chances of maintaining test security that much more.

Stage 3: Administration of the Exam

General Considerations

A license examination process must be administered under standardized and secured conditions. For virtually all large scale testing programs, including the exams under study here, the test development agencies need to be aware of the precautions that must be taken and must administer their exams under these conditions. From what I can tell, both programs administer their tests in ways that conform to professional standards.

The Issue of Testing In Multiple Languages

There is an issue that has arisen with respect to the CALE concerning the fact that the exam is administered in three languages. While there is no documentation that this issue has been raised with respect to the national exam, it is equally applicable to that exam program as well. Therefore, although the discussion here will perhaps have somewhat more of a California focus only because that represents the case where we have more context, an attempt will be made to make the discussion of it relevant to both exams.

The California Code of Regulations governing acupuncture licensure speaks to the languages in which the exam will be administered. Section 1399.441 states:

Examinations shall be administered in English, Chinese, and Korean. Any applicant shall notify the board of the desired language where provided in the Application for Examination/Licensure. Translations and translators, when necessary for other languages, shall be provided in any language for which a translation is formally requested as provided above by a minimum of five

percent (5) of the total number of approved applications. Otherwise, such applicants shall take the examination in one of the languages listed above.

Current Practice

As required, the CALE is routinely offered in English, Mandarin, and Korean. While it is not clear that NCCAOM must also work in multiple languages, it is the case that its exams are also administered in English, Chinese (dialect unspecified), and Korean. Briefly, the process that OER uses in the translation process, which is described in Dr. Ferrel's March 10 response to my inquiries, is as follows. The exam is prepared and finalized in English. It is then given over to their consultant translators who must meet a set of stringent criteria in order to be used in this process. For example, they must be experienced in both languages and both cultures, they must know the subject matter to be translated and must thus be licensed practitioners, and they must possess some basic knowledge of test development and item writing.

The translation of the test is considered to be an adaptation process rather than a literal job of translating the words per say. This adaptation process was described by Dr. Norman Hertz in his testimony to the Little Hoover Commission of September 2, 2003. Dr. Hertz was the chief of OER just prior to Dr. Ferrel, and the above excerpt is from a memo he prepared for dissemination to Marilyn Nielsen (Executive Officer of the Acupuncture Board) dated February 13, 2001. This was an attachment to his written testimony. According to Dr. Hertz:

The OER has been working collaboratively with the Acupuncture Board and its examination contractors for many years in the attempt to prepare examinations that are accurately adapted into Chinese and Korean. We used technology that was psychometrically acceptable at the time. Our process was to translate the examination into Chinese and Korean from English, then back translate in English, and compare the original English with the back translated English. We now know from research studies that this process is flawed; therefore, we have discontinued this practice.

We have begun to use procedures known as "adapting" when taking the English version of the examination and creating Chinese and Korean versions. The adapting process requires that the meaning of the words in the context of the culture be considered when the examinations are created in Chinese and Korean. The adapting process is much more difficult than simply translating words from one language to another. (pp. 2–3 of the February 13, 2001 memo)

NCCAOM also administers its examinations in three languages: English, Chinese, and Korean. What I know of their process is documented in a transcription of an Acupuncture Board meeting which took place on November 18, 2002. Christina Herlihy, the CEO of NCCAOM at that time, presented information concerning the national testing program and engaged in a discussion with the Board. A transcription of that meeting was made available to me by the Little Hoover Commission. Here is the relevant portion of that transcription appearing on page 6. Because this was a transcription of oral exchanges from a taped audio

recording of less than perfect clarity, I will use bracketed text to help the flow of the dialog where the transcription is clearly faulty:

Ms. Asplund [:]...I would like to ask though about the adaptation [of the exam] into Chinese and Korean. Our OER has expressed some concerns in the [past] about the [Chinese and Korean versions being] equivalent to the English language exams, and I'm wondering if the NCCAOM translations or adaptations [are valid].

Ms. Herlihy: We have been reviewed by experts [and found] to be adequate and equal to the standards that we are required to [adhere] to or meet. Our exams are developed in English and all items are developed in [response] to our exam content and outline. The process we are involved in after that [process] is to then put all items forward into a panel in Chinese and Korean. Then [we] review [the] translations that are executed by professional certified medical [translators], what every their whole list is those translations are reviewed by panel of Chinese speaker and health practitioners as well. Their comments go back to the translator and then returned to the committee to make sure the copy and questions have adequately been translated. After the exams have been administered there are separate reviews of the performance of the Chinese candidates as well as the Korean candidates. And the decisions are made in the same fashion as they are for the English exam, as to what items survive the review in the administration process. In other words...there is a separate analysis conducted for the Chinese exam, the English exam as well as the Korean exam. So according to that, the descriptions to our accrediting board our peers...as deemed that as an adequate response to the challenges that we face.

The California translation or adaptation process is state of the art, providing for as good a language conversion process as can currently and reasonably be achieved. The NCCAOM process, while not quite as comprehensive, probably has enough safeguards in it to ensure that the Chinese and Korean versions appear to be testing the same content in approximately the same way that is being done in the English version.

Limitations of Current Practice

The reason for going through an elaborate process of translation or adaptation is to make sure that each language group is being administered the “same” test. Obviously, OER and NCCAOM need to accomplish these linguistic transformations. But there is more to testing the proposition that the tests are essentially the same than just doing good translations. There are statistical analysis strategies that can be applied to the data to address this issue, and both test developers already have enough data to perform at least some of these analyses.

One strategy that can easily be applied to the data is exploratory factor analysis or principle components analysis. If the three language groups are being given the same test, then the internal (factor or component) structure should be comparable. Thus, performing separate analyses on each language groups responses should yield factor structures that reasonably resemble each other. A somewhat related strategy is to perform a confirmatory

factor analysis using the results of, say, those who took the test in English as the basis of the structure. That structure can then be the hypothesized structure for the other two groups, and the analysis can provide many goodness of fit indexes to gauge how well the hypothesized model of the English test group fits the data of the test groups who took the exam in Chinese or Korean. Finally, it is possible to examine the patterns of relationships in a correlation matrix composed of subtest scores by language group.

These statistical strategies, at least the first and the last, are well within the means of both agencies. Such analyses probably should have been done in the past but should, in my opinion, certainly be done in the near future.

The Challenge to Current Practice

Concerns With Current Practice

According to the testimony of Dr. Hertz to the Little Hoover Commission, under his leadership OER has taken a consistent stand since May of 2000 against the use of acupuncture licensing tests in any language but English. As he stated on page 3 of his testimony:

The greatest threat to validity of the current California Acupuncture Licensing Examination (CALE) is that the examination is offered in multiple languages...OER was not successful in my efforts to persuade the Board that the effects from offering the examination in three languages posed a threat to [the] public's health, safety, and welfare; a threat to preventing acupuncture from reaching a "profession" status; and most importantly, the presentation of an examination that is not fair because of its inherent unreliability leading to an examination with questionable evidence of validity.

Analysis of the Concerns With Current Practice

Dr. Hertz's first listed threat to validity is not a technical testing issue and will therefore not be directly addressed. This first threat listed concerns the public's health, safety, and welfare. It has to do with two major issues: integrating acupuncture into the mainstream medical culture (e.g., communicating with other medical practitioners, seeking insurance reimbursement, utilizing new technology) and conforming to reporting requirements concerning possible abuse to children, elders, and others. The argument is that all of these activities are governed by policies and procedures that are ordinarily written and conducted in English. This is a policy issue and is therefore beyond the scope of the present review.

The second threat to validity mentioned in the excerpt deals with acupuncture reaching "professional" status. This has an indirect bearing on the test development process and will be briefly addressed. The argument here is, essentially, that using three languages in testing maintains or accentuates differences in the profession whereas using a single language would serve to help unify the profession. Dr. Hertz stated on page 2 of his February 13, 2001 memo to Marilyn Nielsen:

The occupational analysis that underlies examination development is predicated on the concept that the profession is unified. Consequently, the examination should measure the competencies required for practice and should not attempt to accommodate subjective differences among language groups. Continuing to test in multiple languages imposes a difference that does not exist in practice and adds a difficulty to the examination that is artificial.

In the California occupational analysis conducted by OER, three demographic questions were asked about language: the primary language of the practitioner, which languages the practitioner could speak, and the primary language of the practitioners' patients. Of all the respondents, 160 (about 57%), 56 (about 20%), and 52 (about 18%) indicated that English, Korean, and Chinese was their primary language, respectively. The occupational analysis report does not indicate that there were any differences in the evaluation of the tasks or the knowledge elements among these three groups of practitioners. Since the information existed within their data file, and since their sample size would allow them to explore the issue of potential language group differences, one might therefore be inclined to presume either that (a) the analyses were conducted and no differences were observed, or (b) OER did not consider the issue sufficiently threatening to do the analysis in the first place. Whichever is true, it would appear that even in 2003, at the time of the occupational analysis, the profession was sufficiently unified to have OER use a single job analysis to define the profession despite the fact that multiple language testing had been going on for several years. Thus, the idea that professional practice is different across language groups is not supported by any data analysis that could have rather easily been performed, and therefore, in my opinion, this concern does not comprise a significant threat to the validity of the test.

Dr. Hertz' third mentioned validity threat is that the different language versions, presumably the Mandarin and Korean versions, are not reliable. He stated the following:

...most importantly, the presentation of an examination that is not fair because of its inherent unreliability leading to an examination with questionable evidence of validity.

This is characterized as the most important of the three threats and for good reason. Reliability is a prerequisite to validity. Validity essentially means that you can draw a certain inference from a test score, in this case that it reflects the level of knowledge that candidates demonstrate. But the test score must be a stable estimate of that level knowledge in order for the inference to have any validity, and a reliability index is an indicator of a score's stability.

The reliability of an examination is a part of its quality assessment, a stage of the examination process that is the subject of the next section. As part of that analysis, and as a way to directly address the potential threat to validity raised by Dr. Hertz, the reliability of the test scores based on all candidates as well as the reliability of the different language versions, will be considered.

Recommendations

Since the data exist within the archives of OER, it seems reasonable to actually perform the statistical analyses looking at potential language group differences in the occupational analysis. Each of the three language demographic items can be recoded to create language groups of decent enough size to run some exploratory group differences analyses on the task and knowledge elements ratings. While it certainly would not constitute definitive evidence for either position, it would go a long way to beginning to resolve this issue on an empirical, data-oriented basis.

Stage 4: Quality Assessment

The Context For Performing a Quality Assessment

Once the examination is administered, it is imperative that its quality is assessed before test scores are reported to candidates and certainly before any decisions are made about candidate performance. That is, until such time as the test developer knows that the test scores can be “trusted,” the test itself should remain in a state of limbo. This protects everyone concerned: the candidates themselves, the agency whose test it is, and, ultimately, the public who assume that the agency is making good decisions about who should be licensed.

By the time the quality assessment stage is reached, a high quality examination process has already succeeded in achieving certain ends (and a process of poor quality has already doomed the test to failure). Assuming a high quality test development process, the following goals have already been met: (a) the content that is tested proportionally represents the important portions of the occupation as shown by the job analysis which itself is based on a representative sample of the practitioner population; (b) the test items have been competently and appropriately written by trained item writers; and (c) the test has presumably been administered in a secured and professional manner and the responses of the candidates have been securely conveyed to the agency for scoring.

At this point, the examination is subjected to a quality assessment. This assessment, although based on the data of candidate responses, does not evaluate the candidates but rather targets the test itself. If the test does not meet relatively recognizable standards, it either goes through a post hoc revision process (e.g., removing faulty items, re-keying an item whose supposedly correct response was actually wrong) or, under rare circumstances where the test cannot be “fixed,” it is mercifully destroyed.

The quality assessment is performed in conjunction with the scoring of the exams. This assessment is, fundamentally, a statistical enterprise conducted through computerized analysis of the test data. Although there are more modern ways to analyze information, both OER and NCCAOM use the traditional strategy known as “classical item analysis.” For their quality assessments in the context of licensure examinations, this classical approach is perfectly adequate.

Within this classical analysis strategy, two levels of statistical description are used. At a global level, the test is studied as a whole. From this perspective, we look at descriptive statistics such as the mean, the standard deviation, the standard error of measurement, and the reliability of the scores on the test. At the more microscopic level, we look at individual item performance. This more microscopic analysis focuses on item statistics such as item difficulty levels and item-total correlations.

The Structure Of This Section Of the Report

We have organized the results of the quality assessment of these examinations into a set of tables with accompanying narration. The numerical information will therefore be contained in discrete and bounded locations. We structured the tables after the material produced by OER since it offered the most complete information. The statistical information supplied by NCCAOM was then placed in the corresponding row of the table but, since there was less information on the national exams, many rows of its tables remain blank.

Because we will need to deal with a fair amount of statistical information, we will treat each statistic or related set of statistics in a separate subsection. Recognizing that many readers will have less than working familiarity with these statistics, we will attempt to (a) explain in a minimalist fashion what the statistics represent rather than fully explicating their nature and nuances, (b) describe and compare the California and national tests directly on that statistic or set of statistics, and (c) evaluate the results in straightforward English.

Central Tendency Statistics For the California and NCCAOM Samples as a Whole

Explanation of Central Tendency Measures

Central tendency measures summarize the typical examination scores of the candidates. Of interest here are the statistics representing the number of correct responses and the proportion of items answered correctly.

In a multiple choice test items are scored 1 when candidates answer them correctly and 0 when they do not answer them right. When all of the values of 1 are added together we have the total test score for a candidate. Averaging these total scores gives us the sample mean, shown on the third line of Table 6 through Table 11. For the California January 2003 test, for example, the total sample of 623 candidates averaged around 116 items correct of the 175 that were scored.

The raw scores are useful to know but also it helps to know what they indicate in terms of percentage correct. This statistic is called the “Mean P” value by OER and is shown on the second row from the bottom in the tables. These values are given in decimal form and can be understood as the proportion of the items answered correctly, on average, for the sample. The 116 items correct for the California January 2003 test, for example, corresponds to about 66% correct. More will be said in relation to these P values later.

The median, shown in the 10th row of Tables 6 through 11, is the value that is in the middle of the distribution when the scores are arranged in order. If the shape of the test scores approximates a normal distribution, these values will be very close to their respective means.

Description and Comparison of California and NCCAOM Results Regarding Central Tendency

Since the means are based on different numbers of items, direct comparisons of the various exams are not possible. Specifically, the CALE contained 175 items in both of the testing cycles for which we had results. NCCAOM's March 2002 acupuncture test contained 180 items but their other acupuncture and Chinese Herbology tests contained 200 items.

The NCCAOM tests, however, are associated with a certain vagueness concerning the item count. As documented in Attachment F, NCCAOM places their 20 newly created items on the test and scores them. These items will be counted toward the total test score unless the item performs in an undesirable way (items with difficulty indexes in the .99 region, for example, may or may not be scored). In reviewing the NCCAOM tests, we therefore have no way of knowing from the information supplied how many of the items appearing on the test were actually counted and thus contributed to (a) the test score of candidates and (b) the statistical analyses summarized below.

In dealing with average performance, using the mean proportion of correct responses allows us to talk about and compare the tests on the same metric. This statistic is the mean P value and tells us multiple elements simultaneously among which are the: (a) average proportion of items that candidates correctly answered; (b) probability of candidates selecting the correct answer; (c) average difficulty of the test.

For the full 2003 California sample, the mean difficulty levels for January (shown in Table 6) and August (shown in Table 7) were around .66 and .70, respectively. This tells us that candidates were correctly responding to about two thirds of the test questions. Corresponding values for NCCAOM's tests are shown in separately. Tables 8 and 9 report the summaries of the acupuncture modules for the last couple of years. Difficulty levels are in the low to middle .8s indicating that candidates were answering about 80% to 85% of the items correctly. The Chinese Herbology modules ranged in difficulty between .76 and .82, thus proving to be just slightly more difficult than the NCCAOM acupuncture tests.

Evaluation of the Results Regarding Central Tendency

Items are intended to help tell the difference between test takers who are more and less competent in their knowledge of the profession. One measure of an items' ability to carry out this mission is its variance, which is a function of its item difficulty level. Generally speaking, and all else equal, items with greater variance will do a better job in distinguishing candidates. Item variance is maximized at a difficulty level of .50; as difficulty levels shift away from this toward either end of the scale, item variance decreases. The NCCAOM

items, having lower variances on average, might seem on the surface to be a bit weaker than the California items, at least based on the item difficulty index. As it turns out, that is not actually the case based on the point-biserial correlation results.

Central Tendency Statistics For the California Language Groups

Since there have been some concerns expressed about the California license examinations being offered in multiple languages, it is useful to address this issue when the statistical results permit us to do so. It can be seen in Tables 6 and 7 that the means for English, Mandarin, and Korean versions of the test for both January and August 2003 differ slightly. One might think that while the English and Mandarin groups achieved scores that were pretty close to each other, the Korean sample might have done worse than the other two in January and better than the other two in August.

It is possible to determine if these differences represent chance variations and that the means are therefore essentially the same or whether the differences are actually statistically significant different from each other. The large sample sizes (in the context of comparing these means) probably give us more statistical power than we would need or even want, but we engaged in the analysis anyway. Because the mean, standard deviation, and group sizes were already computed by OER and placed in their summary tables, it was convenient to run a series of *t* tests comparing every different pair of means within a given cycle. That is, we compared English to Mandarin, English to Korean, and Mandarin to Korean for each cycle. With a Bonferroni correction to our default .05 alpha level, none of the comparisons resulted in a statistically significant *t* value. One would therefore conclude that the minor fluctuations among the language groups would be expected to occur on the basis of chance and do not reflect meaningful differences among the groups. That is, the groups performed comparably to each other on average with the differences among the means within the margin of error.

Variability Statistics For the California and NCCAOM Samples as a Whole

Variability measures summarize the dispersion of the test scores. As shown in Tables 6 through 11, the standard deviation is the only measure of variability computed by both test analysis groups, and this describes the scattering of test scores around the sample mean. There is little to say about this statistic other than all of the reported values make sense given their respective means.

California computes several other useful variability measures that together more completely describe the score distribution. Since we cannot compare the CALE to the NCCAOM exams, there is little to be gained by describing California's results in detail. Of these other measures the most important of them are kurtosis and skewness. Briefly, the January test yielded pretty much of a normal distribution whereas the August test distribution tended to be a bit platykurtic (flattened) and somewhat negatively skewed (test takers are a little more bunched up at the higher scores).

Reliability Statistics For the California and NCCAOM Samples as a Whole

The Concept of Reliability

Reliability is used in the field of testing in a way similar to its use in everyday language. When we say that we can rely on some given information we mean that it sufficiently informative to use as a basis of a decision. In a general sense, that works in testing as well. We use the test score to make a decision about whether or not candidates are awarded a license; we thus rely on the test score to be useful in that capacity.

The issue of reliability comes in to how much we can rely on the test scores as a measure of professional knowledge because, as described in a moment, we understand that the test score is only an estimate of what candidates know—there is always some uncertainty (or measurement error) embedded in that information. Reliability is a gauge of the amount of such uncertainty contained in the test scores. Reliability indexes typically range between 0 and 1. Values around 0 indicate that there is so much uncertainty in the measurement that we simply cannot rely on the test score to tell us anything of worth; values around 1 tell us that the uncertainty is minimal and that the test score is a useable measure of whatever it is that the test was validly assessing.

The degree to which reliability impacts the use of license test scores depends, to a certain extent, on the passing rate. If virtually everyone passes or if virtually everyone fails the test regardless of the test score they achieved, then lower reliability is less of a concern since everyone's fate is the same regardless of their test performance. Thus, even if there is more measurement error than we would prefer, it would make little practical difference since the test scores of the candidates appear to be unrelated to the decision to pass or fail them.

But as we approach an even division of those who passed or failed—as we see a more even division of scores above and below the pass point—we are increasingly reliant on the information communicated by the test score, and the reliability of those scores becomes an increasingly critical piece of information. Now one possible consequence of having increasingly lower reliability (increasingly more measurement error) is that more of those candidates who possessed considerable knowledge could have scored in the failing range and more of those candidates who knew relatively little could have scored high enough to pass. These decision errors, false negatives and false positives, respectively, could have been minimized, all else equal, if the reliability of the test scores would have been higher.

Measurement Error and Reliability

Measurement error, the fundamental concept driving this discussion, is explicitly studied in at least two ways in the test analysis: by estimating the reliability of the scores and by applying the standard error of measurement to the scores. We will look at reliability here and the standard error of measurement in the following subsection.

Reliability can be assessed in several different ways. We have discussed two of these already in the context of analyzing the responses to the job analysis questionnaire. Internal consistency using coefficient alpha addresses the issue of how consistently the practitioners responded to the items, and it was stated earlier that this reliability index was less useful for evaluating the job analysis questionnaire responses than the second approach to reliability. This second approach was to assess rater reliability using the intraclass correlation which addresses the issue of the degree to which the respondents are “in synch” with each other.

Now that we are in the test analysis stage, we are not interested in respondent synchrony. The respondents are now the licensing candidates who may or may not know much about their profession, not the SMEs who are already expert practitioners, and we are perfectly prepared to find that the candidates may not at all be “in synch” with each other—some of them may possess considerable knowledge and answer the exam questions in one way whereas others may possess little knowledge and thus answer the exam questions in quite a different manner.

On the other hand, we are now very concerned with the relationships among the items. If the test questions are reasonable gauges of knowledge then practitioners should be responding to these items in a reasonably consistent manner. For example, if X, Y, and Z are related pieces of knowledge then practitioners who have answered questions on X correctly should also tend to answer questions on Y and Z correctly as well. Coefficient alpha can assess this proposition by indexing the degree of consistency contained in the test data.

As indicated earlier in this report, although it is possible for coefficient alpha to take on negative values it ideally ranges between .00 and 1.00. But we are not interested in the full range of this index—we would expect that licensing tests of the length used here (this statistic is sensitive to test length) should be at least in the middle to high .8s to meet acceptable standards. California hit this mark in its January 2003 test and reached into the low .9s in August of that year. NCCAOM was consistently in the low .9s throughout the last two years. We can therefore say that the scores on both tests would therefore be considered to be quite reliable.

Measurement Error and the Standard Error of Measurement

Measuring the knowledge of acupuncture that candidates possess must of necessity entail a certain amount of imprecision. This is true in physics, chemistry, psychology, and licensure testing. Within quantum mechanics, for example, the location of a particular electron circling the nucleus of an atom cannot be precisely known but is rather described by a probability function.

The imprecision in testing, for example, the uncertainty of identifying a candidate’s exact level of acupuncture knowledge, is called measurement error by scientists and statisticians. It is meant to represent the fact that scientists would never be able to exactly reproduce the same measurement outcome were they to repeat their measurement process an indefinite number of times due to the interaction between test item and practitioner inconsistency across the test versions. Such uncertainty is to be expected partly because we are human and partly because that seems to be the way the universe is built. The term “error” is used to describe this inconsistency but it is not meant to suggest that mistakes have been made in the

measurement procedures. Perhaps a better term to use in representing this concept is “expected fluctuations” but tradition has decreed that we must use the term “error” instead.

Reliability is one way to index the degree of measurement error in a test; higher reliability indexes greater precision and less error. Test reliability can be translated into a test score value that concretely applies it to candidate performance. The statistic with which this is done is called the standard error of measurement. Both test analysis groups report this statistic in their summary. This outcome of the test analysis appears on Tables 6 through 11 on the row labeled as “SEM.”

The general idea behind this application is as follows. A candidate obtains a particular test score on the license examination, but by definition there is some imprecision (measurement error) associated with that assessment. The candidate might have scored somewhat higher or somewhat lower on any given day despite the fact that they would have possessed the same amount of knowledge. How much higher or lower is told to us by the standard error of measurement. For example, the odds are 2:1 that candidates would likely have scored between plus and minus one standard error unit because of the imprecision in the

To illustrate this, consider a candidate who scored 116 on the January 2003 California test. With the standard error of measurement equal to roughly five and a half test points, we can say that the candidate’s likely performance given the imprecision of the test is really within the range of 110.5 to 121.5. This band, which is relatively narrow but still spans some range of test scores, tells us information that supplements the single test score value. This information can be used in several important ways. For example, the imprecision of a licensing test should be taken into account when a passing score is established.

Reliability For the Different Language Groups in the California Examinations

In the section of this report devoted to Stage 3, the administration of the exam, the issue of testing in different languages was raised. Treatment of the reliability question was postponed to this current section and we can examine it here. As will be recalled, the following concern was raised regarding the California exam being administered in three different languages:

...most importantly, the presentation of an examination that is not fair because of its inherent unreliability leading to an examination with questionable evidence of validity.

Tables 6 and 7 can directly evaluate the degree to which unreliability threatens the validity of the CALE. As we now know, the overall test is very reliable, with reliability coefficients of .891 and .925 for the January and August 2003 exams. When the sample is divided into language groups, the strong reliability holds across the board for both testing cycles. In short, the test does not lose any precision when we look at the English, Mandarin, or Korean versions in isolation. Thus, the serious concerns that were raised regarding a decrement in reliability leading to a loss of validity do not at all appear to have actually materialized.

Table 6
California State Examination
Statistics for January 2003

	CA January ALL^a	CA January English	CA January Mandarin	CA January Korean
Number of Items	175	175	175	175
Number of Examinees	623	383	148	92
Mean	116.130	116.752	116.122	107.717
Variance	289.491	259.726	257.040	365.529
Standard Deviation	17.014	16.116	16.032	19.119
Skewness	-.454	-.446	-.126	-.516
Kurtosis	.257	.163	-.438	.249
Minimum	56	62	73	55
Maximum	156	150	155	153
Median	117	117	116	109
Reliability	.891	.882	.881	.913
SEM	5.627	5.540	5.527	5.626
Mean P ^b	.664	.667 (.6653)	.664 (.6634)	.616 (.6161)
Mean Point-Biserial ^c	.205 (.227)	.196 (.223)	.193 (.230)	.234 (.283)
Max Score (Low)	107	107	106	100
N (Low Group)	178	103	42	26
Min Score (High)	128	127	127	119
N (High Group)	169	115	42	27

^a The “All” column for the January 2003 exam should represent a weighted average of the three language groups since together these groups comprise the total sample for that test. However, the values for “All” are different by a small margin from the weighted average of the three groups. Without access to the raw data, we did not engage in the statistical diagnostic procedures to determine where in the various analyses the discrepancy might be located.

^b This can be understood as the mean proportion of items answered correctly. We have recalculated this statistic for the separate language groups and have obtained a slightly different mean value than the mean value that was originally reported by the State. The value in the parenthesis is the one we calculated.

^c This value was originally called “Mean Item-Total” by the California State. We have recalculated this statistic by a more appropriate method (we squared the *rs*, averaged them, and then took the square root of the average; *r* departs sufficiently from interval measurement so that its values cannot properly be directly averaged) and have obtained a slightly different mean value than the mean value that was originally reported by the State. The value in parenthesis is the one we calculated.

Table 7
California State Examination
Statistics for August 2003

	CA August ALL	CA August English	CA August Mandarin	CA August Korean
Number of Items	175	175	175	175
Number of Examinees	606	341	142	123
Mean	121.640	121.894	117.873	125.285
Variance	405.731	397.761	397.266	407.764
Standard Deviation	20.143	19.94	19.932	20.193
Skewness	-.726	-.641	-.864	-.893
Kurtosis	.694	.460	1.246	.821
Minimum	47	61	47	48
Maximum	162	162	155	162
Median	124	123	120	129
Reliability	.925	.924	.924	.933
SEM	5.517	5.483	5.493	5.244
Mean P	.695	.697	.674	.716
Mean Point Biserial ^a	.251 (.268)	.250 (.270)	.253 (.282)	.273 (.309)
Max Score (Low)	112	113	108	116
N (Low Group)	164	98	39	36
Min Score (High)	135	134	129	141
N (High Group)	166	96	41	34

^a This value was originally called “Mean Item-Total” by the California State. We have recalculated this statistic by a more appropriate method (we squared the *rs*, averaged them, and then took the square root of the average; *r* departs sufficiently from interval measurement so that its values cannot properly be directly averaged) and have obtained a slightly different mean value than the mean value that was originally reported by the State. The value in parenthesis is the one we calculated.

Table 8
NCCAOM Acupuncture Examination
Statistics for 2002

	NCCAOM Acupuncture March 2002	NCCAOM Acupuncture July 2002	NCCAOM Acupuncture November 2002
Number of Items	180	200	200
Number of Examinees	631	568	527
Mean	153.49	165.10	167.17
Variance			
Standard Deviation	18.99	20.71	17.48
Skewness			
Kurtosis			
Minimum	50	61	69
Maximum	179	193	192
Median			
Reliability	.94	.95	.94
SEM	4.50	4.73	4.45
Mean P	.83	.83	.84
Mean Point Biserial ^a	.31	.31	.28
Max Score (Low)			
N (Low Group)			
Min Score (High)			
N (High Group)			

^a This value was originally called “Mean Item Discrimination” by the NCCAOM.

Table 9
NCCAOM Acupuncture Examination
Statistics for 2003

	NCCAOM Acupuncture February 2003	NCCAOM Acupuncture June 2003	NCCAOM Acupuncture October 2003
Number of Items	200	200	200
Number of Examinees	567	636	670
Mean	169.02	160.76	165.13
Variance			
Standard Deviation	17.58	17.59	19.37
Skewness			
Kurtosis			
Minimum	69	60	80
Maximum	196	190	195
Median			
Reliability	.93	.92	.94
SEM	4.54	4.84	4.89
Mean P	.85	.80	.83
Mean Point Biserial ^a	.28	.26	.28
Max Score (Low)			
N (Low Group)			
Min Score (High)			
N (High Group)			

^a This value was originally called “Mean Item Discrimination” by the NCCAOM.

Table 10
NCCAOM Herbology Examination
Statistics for 2002

	NCCAOM Chinese Herbology March 2002	NCCAOM Chinese Herbology July 2002	NCCAOM Chinese Herbology November 2002
Number of Items	200	200	200
Number of Examinees	222	235	228
Mean	147.61	160.87	153.34
Variance			
Standard Deviation	17.83	19.94	17.60
Skewness			
Kurtosis			
Minimum	68	68	90
Maximum	176	192	186
Median			
Reliability	.93	.94	.91
SEM	4.85	5.02	5.26
Mean P	.82	.80	.77
Mean Point Biserial ^a	.27	.27	.23
Max Score (Low)			
N (Low Group)			
Min Score (High)			
N (High Group)			

^a This value was originally called “Mean Item Discrimination” by the NCCAOM.

Table 11
NCCAOM Herbology Examination
Statistics for 2003

	NCCAOM Chinese Herbology February 2003	NCCAOM Chinese Herbology June 2003	NCCAOM Chinese Herbology October 2003
Number of Items	200	200	200
Number of Examinees	172	270	232
Mean	156.98	162.60	152.06
Variance			
Standard Deviation	20.74	18.11	19.41
Skewness			
Kurtosis			
Minimum	93	58	75
Maximum	192	194	185
Median			
Reliability	.94	.93	.93
SEM	5.20	4.93	5.31
Mean P	.78	.81	.76
Mean Point Biserial ^a	.28	.26	.25
Max Score (Low)			
N (Low Group)			
Min Score (High)			
N (High Group)			

^a This value was originally called “Mean Item Discrimination” by the NCCAOM.

Item Difficulty Statistics For the California and NCCAOM Samples as a Whole

Explanation of Item Difficulty

Items on these tests are scored 1 for correct and 0 for incorrect responses. Adding the 1s and 0s across the test questions for each candidate allows us to obtain his or her total score. Averaging the 1s and 0s for each item across all of the candidates allows us to obtain the mean item difficulty value. Note that the focus here is on the test items and not on the candidates.

Item difficulty values can range between 0 and 1. With 1s signifying correct responses, items with more 1s (and thus higher difficulty values) are easier. Easier items have means closer to 1. If few candidates answered the question correctly, that item would be associated mostly with 0s and the mean would be near 0. If almost everyone answered the question correctly, that item would be dominated by 1s and the mean would be near 1. Items which were answered correctly by half the respondents would have an equal number of 0s and 1s and the mean item difficulty value would be .5.

The technical label for this statistic is “item difficulty” but an argument can be made that it should be called “item easiness” since larger values tell us that the item is easier. Items with difficulty values lower than .20 are considered quite difficult and those with difficulty values higher than .80 are considered quite easy. Items more difficult than .10 and easier than .90 are thought of as rather extreme. Within classical item analysis, all else equal, these latter values are considered to be less desirable.

We have already addressed the overall item difficulty value for each test. Tables 12 through 16 present a more microscopic analysis of these statistical results. For each test, we have counted the number of items whose difficulty value falls within a specified range. For example, consider the California January 2003 exam. We can see from Table 12 that 12 items, which represent 6.9% of the total number of items on the test, were associated with item difficulty values between .30 and .39.

Description and Comparison of California and NCCAOM Results Regarding Item Difficulty

The California test results are shown in Table 12. As can be seen, the bulk of items had difficulty values in the .4s through the .8s with a bit of a concentration in the .6s, .7s, and .8s. From a psychometric perspective, if not from a candidate point of view, these California tests would be thought of as moderate in difficulty.

The NCCAOM tests present a somewhat different picture. Both of the modules for which we had information were shifted toward the easier end of the continuum, especially the acupuncture module. The items in this module tend to have difficulty values in the .8s and .9s with the .9 range always more heavily populated. To have at times upwards of half or close to half of your 200 items answered correctly by 90% of your test takers suggests from a psychometric perspective that the test is very easy.

The Chinese Herbology module appears to be somewhat more challenging to candidates than their acupuncture module, but still the difficulty levels are skewed toward the high end of the continuum. Here, the general concentration seems to be in the .7s, .8s, and .9 but the .8s and .9s are still more populated than the .7 range.

These NCCAOM results may be a tiny bit “fuzzy,” but not necessarily to an alarming extent. As was mentioned in connection with the earlier discussion of central tendency measures, my communication with NCCAOM (see Attachment F) revealed that some (but apparently not all) of the items with difficulty values of .99 or 1.00 may be removed from the scoring protocol used to evaluate candidate exams. Despite the fact that information for 200 items is shown in the tables, the mean difficulty level of the items reported by NCCAOM is based on less than 200 items (our own calculations for all 200 items yielded a different value from that reported by NCCAOM). However, there was no documentation indicating which items were removed. Without such documentation, it is impossible to determine exactly what was done and how well it was done.

Evaluation of the Results Regarding Item Difficulty

Within classical item analysis, items at the extremes of difficulty (either very easy or very difficult) are viewed as supplying a limited amount of information (their variance, a measure of information, is very restricted) and are therefore considered to be less desirable. From that perspective, the California tests would be judged to be quite acceptable but the NCCAOM exams, especially the acupuncture module, might be the cause of some concern.

However, two additional elements need to be factored in to such an evaluation: the discriminating ability of the items and, in a licensing exam context, where the pass points are set. Easy items may still be able to distinguish among candidates. To the extent that they can do so, this feature can offset the liabilities of using extremely easy items. Also, if the pass point is set near the region of the average item difficulty, the items could actually be serving the useful purpose of helping the test developers to tell apart those with minimal competency from those who lack it.

One very effective way to examine these two item features is to plot an item characteristic curve for each item. This analysis can be accomplished in the classical approach used here by both test development groups (using ranges of total test score on the x axis instead of \square). The resulting graphical summary can readily allow test developers to see how well the item discriminates among test takers of different competency levels while at the same time identifying which competency levels the item is best able to differentiate. Unfortunately, neither testing group made use of this form of analysis. The issues concerning items being able to differentiate test takers must therefore be explored a bit more indirectly by evaluating the point-biserial correlations associated with each variable. The next section of the report covers this topic.

Table 12
Frequency Tabulation of Item Difficulty
Values for 2003 State of California
License Examinations

Difficulty Values	CA State January 2003		CA State August 2003	
	Number of Items	% of Total	Number of Items	% of Total
.00 - .09	0	0%	0	0%
.10 - .19	0	0%	0	0%
.20 - .29	5	2.9%	2	1.1%
.30 - .39	12	6.9%	8	4.6%
.40 - .49	18	10.3%	12	6.9%
.50 - .59	23	13.1%	21	12.0%
.60 - .69	28	16.0%	39	22.3%
.70 - .79	38	21.7%	39	22.3%
.80 - .89	35	20.0%	40	22.9%
.90 - .99	16	9.1%	14	8.0%
1.00	0	0%	0	0%
	Overall Mean = .664		Overall Mean = .695	
Total	175	100%	175	100%

Table 13
Frequency Tabulation of Item Difficulty
Values for 2002 NCCAOM
Acupuncture Examinations

Difficulty Values	NCCAOM March 2002		NCCAOM July 2002		NCCAOM November 2002	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
.00 - .09	0	0%	0	0%	0	0%
.10 - .19	0	0%	0	0%	0	0%
.20 - .29	0	0%	0	0%	1	0.5%
.30 - .39	0	0%	2	1.0%	1	0.5%
.40 - .49	6	3.0%	4	2.0%	1	0.5%
.50 - .59	5	2.5%	9	4.5%	6	3.0%
.60 - .69	11	5.5%	7	3.5%	9	4.5%
.70 - .79	24	12.0%	29	14.5%	25	12.5%
.80 - .89	58	29.0%	73	36.5%	59	29.5%
.90 - .99	92	46.0%	76	38.0%	96	48.0%
1.00	4	2.0%	0	0%	2	1.0%
	Overall Mean = .83		Overall Mean = .83		Overall Mean = .84	
Total	200	100%	200	100%	200	100%

Table 14
Frequency Tabulation of Item Difficulty
Values for 2003 NCCAOM
Acupuncture Examinations

Difficulty Values	NCCAOM February 2003		NCCAOM June 2003		NCCAOM October 2003	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
.00 - .09	1	0.5%	0	0%	0	0%
.10 - .19	2	1.0%	0	0%	0	0%
.20 - .29	1	0.5%	2	1.0%	0	0%
.30 - .39	0	0%	6	3.0%	1	0.5%
.40 - .49	1	0.5%	3	1.5%	5	2.5%
.50 - .59	5	2.5%	7	3.5%	5	2.5%
.60 - .69	12	6.0%	16	8.0%	17	8.5%
.70 - .79	28	14.0%	29	14.5%	32	16.0%
.80 - .89	50	25.0%	55	27.5%	65	32.5%
.90 - .99	100	50.0%	82	41.0%	75	37.5%
1.00	0	0%	0	0%	0	0%
	Overall Mean = .85		Overall Mean = .80		Overall Mean = .83	
Total	200	100%	200	100%	200	100%

Table 15
Frequency Tabulation of Item Difficulty
Values for 2002 NCCAOM
Chinese Herbology Examinations

Difficulty Values	NCCAOM March 2002		NCCAOM July 2002		NCCAOM November 2002	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
.00 - .09	0	0%	1	0.5%	2	1.0%
.10 - .19	0	0%	0	0%	0	0%
.20 - .29	0	0%	0	0%	1	0.5%
.30 - .39	1	0.5%	1	0.5%	5	2.5%
.40 - .49	5	2.5%	6	3.0%	10	5.0%
.50 - .59	13	6.5%	12	6.0%	17	8.5%
.60 - .69	20	10.0%	18	9.0%	24	12.0%
.70 - .79	32	16.0%	35	17.5%	28	14.0%
.80 - .89	52	26.0%	61	30.5%	58	29.0%
.90 - .99	73	36.5%	65	32.5%	55	27.5%
1.00	4	2.0%	1	0.5%	0	0%
	Overall Mean = .82		Overall Mean = .80		Overall Mean = .77	
Total	200	100%	200	100%	200	100%

Table 16
Frequency Tabulation of Item Difficulty
Values for 2003 NCCAOM
Chinese Herbology Examinations

Difficulty Values	NCCAOM February 2003		NCCAOM June 2003		NCCAOM October 2003	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
.00 - .09	0	0%	0	0%	1	0.5%
.10 - .19	0	0%	1	0.5%	0	0%
.20 - .29	0	0%	0	0%	2	1.0%
.30 - .39	2	1.0%	0	0%	6	3.0%
.40 - .49	13	6.5%	8	4.0%	10	5.0%
.50 - .59	14	7.0%	11	5.5%	17	8.5%
.60 - .69	17	8.5%	15	7.5%	23	11.5%
.70 - .79	39	19.5%	31	15.5%	37	18.5%
.80 - .89	63	31.5%	62	31.0%	50	25.0%
.90 - .99	50	25.0%	71	35.5%	51	25.5%
1.00	2	1.0%	1	0.5%	3	1.5%
	Overall Mean = .78		Overall Mean = .81		Overall Mean = .76	
Total	200	100%	200	100%	200	100%

Point-Biserial Statistics For the California and NCCAOM Samples as a Whole

Explanation of the Point-Biserial Correlation

Each item on a test taps into the knowledge of one content (knowledge) element underlying the profession. To assess candidates' knowledge of the entire profession, the test questions are aggregated together to yield a total test score. Although we tend to think of the test score as a single entity, it is really a composite of the test questions on which it is based.

Up to some point, the analogy that can be used here is a brick wall where the individual bricks are the test items and the wall as an entity is the total test score. This analogy is useful in helping us remember that each of the bricks must be placed in a position that supports the integrity of the wall. Specifically, bricks that are well placed add strength to the wall and bricks that are poorly placed weaken the wall. In a quality assessment of the brick wall, we would want to spot poorly placed bricks and replace them with better positioned ones.

What is true for the wall in the above analogy is true for a licensing test. Each individual item needs to "support" the test as a whole for the test to have integrity. "Support" in this instance means that earning the point value for the item (achieving the 1 point for the item by answering it correctly) means that candidates have more knowledge of the profession than those who do not get the right answer, just as higher test scores mean that candidates have demonstrated greater knowledge than those who earn a lower score.

The statistic that assesses this meaning of "support" is the corrected point-biserial correlation. Essentially, this is the correlation of the item (which is scored 1 or 0) with the

total score of the remaining test items. We recognize that the item supports the test when it yields a positive correlation. As a very general guide, acceptable support is indexed by correlations in the teens or sometimes slightly below, good support is indexed by correlations in the .2s and .3s, and very good support is indexed by correlations in the .4s or higher.

Corrected point-biserial (item-total) correlations around zero do no real damage to the test but do not add any support to the test as a whole. Negative corrected item-total correlations are danger signals; items with negative point-biserial correlations lower the reliability of the test and make the test much less valid. Such items need to be either rehabilitated (re-keyed if an incorrect answer was programmed as the right choice) or purged from the test prior to final scoring.

In classical item analysis where item characteristic curves are not produced, the corrected point-biserial correlations are one of the most important of the single item summary statistics on which test developers rely when they perform their quality assessment of the test. High positive correlations for this statistic can even override concerns about extreme item difficulty values.

Description and Comparison of California and NCCAOM Results Regarding Point-Biserial Correlations

There was no documentation by either test group on whether or not the point-biserial correlations they computed were of the corrected variety, although most statistical programs specifically designed to perform test analysis do calculate this version. With the relatively large number of items involved on these tests, the failure to use the corrected version would have a minimal impact on the results in any case. In the following description of the results, we will call coefficients point-biserial correlations and avoid the term “corrected” so as not to ascribe more to the statistic than what the agencies themselves reported to us.

Tables 17 through 21 present the distribution of the point-biserial correlations grouped by intervals of .10. Table 17 shows the California results for the two tests administered in 2003. The interval containing the most correlations is the one for the .2s. Generally, about two thirds of the values for the January test and about three quarters of the values for the August test would have items considered to be good or better by the standards described above.

The results for the NCCAOM acupuncture exams are shown in Tables 18 and 19. For 2002, the interval containing the highest frequency of point-biserial correlations is the one for the .3s while in 2003 it is mostly the .2s. Regardless, about 80% of the 2002 items and about 85% of the 2003 acupuncture items had point-biserial correlations that would be considered to be good. The NCCAOM Chinese Herbology exams are shown in Tables 20 and 21. They present a similar if not quite as strong picture. About two thirds of the 2002 items and somewhat more than two thirds of the 2003 items would be considered to have yielded good item-total correlations.

**Evaluation of the Results
Regarding the Point-
Biserial Correlations**

Both the state and the national examination processes produced good quality items. This was the case despite the fact the NCCAOM items, especially the acupuncture items, were relatively easy for candidates. Thus, generally speaking, the items on both tests for the most part are sensitive to the presumed knowledge level of the candidates; those with more knowledge of their subject matter, at least as gauged by their total test performance, are more likely to achieve the correct answer on the test questions than those with less knowledge of the profession. This is exactly the way in which we would expect the items to behave in a high quality test.

**Table 17
Frequency Tabulation of Item Point-Biserial
Correlations for 2003 State of California
License Examinations**

Point-Biserial	CA State January 2003		CA State August 2003	
	Number of Items	% of Total	Number of Items	% of Total
Less than .00	3	1.7%	1	0.6%
.00 - .09	23	13.1%	7	4.0%
.10 - .19	41	23.4%	41	23.4%
.20 - .29	79	45.1%	69	39.4%
.30 - .39	27	15.4%	47	26.9%
.40 - .49	2	1.1%	10	5.7%
.50 - .59	0	0%	0	0%
.60 - .69	0	0%	0	0%
.70 - .79	0	0%	0	0%
.80 - .89	0	0%	0	0%
.90 - .99	0	0%	0	0%
1.00	0	0%	0	0%
	Overall Mean = .205 (.227)		Overall Mean = .251 (.268)	
Total	175	100%	175	100%

Table 18
Frequency Tabulation of Item Point-Biserial
Correlations for 2002 NCCAOM
Acupuncture Examinations

Point-Biserial	NCCAOM March 2002		NCCAOM July 2002		NCCAOM November 2002	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
Less than .00	0	0%	1	0.5%	0	0%
.00 - .09	5	2.5%	2	1.0%	7	3.5%
.10 - .19	19	9.5%	24	12.0%	35	17.5%
.20 - .29	44	20.0%	58	29.0%	60	30.0%
.30 - .39	76	38.0%	67	33.5%	71	35.5%
.40 - .49	47	23.5%	38	19.0%	25	12.5%
.50 - .59	9	4.5%	8	4.0%	2	1.0%
.60 - .69	0	0%	0	0%	0	0%
.70 - .79	0	0%	0	0%	0	0%
.80 - .89	0	0%	0	0%	0	0%
.90 - .99	0	0%	0	0%	0	0%
1.00	0	0%	0	0%	0	0%
Missing Data	0	0	2	1.0%	0	0%
	Overall Mean = .31		Overall Mean = .31		Overall Mean = .28	
Total	200	100%	200	100%	200	100%

Table 19
Frequency Tabulation of Item Point-Biserial
Correlations for 2003 NCCAOM
Acupuncture Examinations

Point-Biserial	NCCAOM February 2003		NCCAOM June 2003		NCCAOM October 2003	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
Less than .00	3	1.5%	2	1.0%	0	0%
.00 - .09	5	2.5%	8	4.0%	5	2.5%
.10 - .19	34	17.0%	44	22.0%	31	15.5%
.20 - .29	64	32.0%	67	33.5%	70	35.0%
.30 - .39	67	33.5%	58	29.0%	68	34.0%
.40 - .49	26	13.0%	21	10.5%	25	12.5%
.50 - .59	1	0.5%	0	0%	0	0%
.60 - .69	0	0%	0	0%	0	0%
.70 - .79	0	0%	0	0%	0	0%
.80 - .89	0	0%	0	0%	0	0%
.90 - .99	0	0%	0	0%	0	0%
1.00	0	0%	0	0%	0	0%
Missing Data	0	0%	0	0%	1	0.5%
	Overall Mean = .28		Overall Mean = .26		Overall Mean = .28	
Total	200	100%	200	100%	200	100%

Table 20
Frequency Tabulation of Item Point-Biserial
Correlations for 2002 NCCAOM
Chinese Herbology Examinations

Point-Biserial	NCCAOM March 2002		NCCAOM July 2002		NCCAOM November 2002	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
Less than .00	4	2.0%	4	2.0%	4	2.0%
.00 - .09	18	9.0%	14	7.0%	21	10.5%
.10 - .19	33	16.5%	34	17.0%	42	21.0%
.20 - .29	59	29.5%	58	29.0%	70	35.0%
.30 - .39	51	25.5%	51	25.5%	49	24.5%
.40 - .49	28	14.0%	35	17.5%	14	7.0%
.50 - .59	7	3.5%	4	2.0%	0	0%
.60 - .69	0	0%	0	0%	0	0%
.70 - .79	0	0%	0	0%	0	0%
.80 - .89	0	0%	0	0%	0	0%
.90 - .99	0	0%	0	0%	0	0%
1.00	0	0%	0	0%	0	0%
Missing Data	0	0	0	0%	0	0%
	Overall Mean = .27		Overall Mean = .27		Overall Mean = .23	
Total	200	100%	200	100%	200	100%

Table 21
Frequency Tabulation of Item Point-Biserial
Correlations for 2003 NCCAOM
Chinese Herbology Examinations

Point-Biserial	NCCAOM February 2003		NCCAOM June 2003		NCCAOM October 2003	
	Number of Items	% of Total	Number of Items	% of Total	Number of Items	% of Total
Less than .00	2	1.0%	1	0.5%	9	4.5%
.00 - .09	14	7.0%	11	5.5%	14	7.0%
.10 - .19	34	17.0%	48	24.0%	31	15.5%
.20 - .29	64	32.0%	61	30.5%	64	32.0%
.30 - .39	49	24.5%	55	27.5%	57	28.5%
.40 - .49	28	14.0%	22	11.0%	22	11.0%
.50 - .59	8	4.0%	1	0.5%	3	1.5%
.60 - .69	1	0.5%	0	0%	0	0%
.70 - .79	0	0%	0	0%	0	0%
.80 - .89	0	0%	0	0%	0	0%
.90 - .99	0	0%	0	0%	0	0%
1.00	0	0%	0	0%	0	0%
Missing Data	0	0%	1	0.5%	0	0%
	Overall Mean = .28		Overall Mean = .26		Overall Mean = .25	
Total	200	100%	200	100%	200	100%

Stage 5: Pass Point Analysis

Establishing a Pass Point

It is common in the arena of licensure and certification to establish a passing point for the examination using some variant of an Angoff procedure, and both OER and NCCAOM utilize essentially similar standard setting process. Introduced in 1971, this procedure calls for a panel of SMEs to review each test question and to estimate the proportion of candidates meeting a “minimal competence” criterion who would, in their judgment, answer the question correctly. Ratings for each SME are summed and the sums of all SMEs are averaged to obtain a consensus proportion representing the test. That proportion then becomes the passing score for the test (more likely, the first approximation to the passing score since it should be modified by taking into account the standard error of measurement). The result of the Angoff procedure is interpreted as indicating the minimum score candidates can achieve to indicate to the license or certification agency that they possess the minimal competency needed to be licensed or certified.

The Angoff method and its variants may not be perfect procedures but they can produce valid pass points when the testing agencies follow an appropriate set of procedures. For example, the panel needs to consist of knowledgeable individuals who represent the diversity of the profession. The panel must also be large enough so that the ratings of a single SME do not distort the judgments of the majority. Training needs to be sufficiently complete for the panel members to reach a clear and uniform understanding of what “minimal competency” means. The performance of the panel needs to be monitored and feedback provided at least in the early phases of their work. Ideally, there should be some mechanisms built in to the process to ensure that the panel members do not reach substantially different findings.

The California Process

Based on the material supplied by Dr. Ferrel in her March 10 letter, OER uses 8 to 10 practitioners in their panels with at least half who have been licensed for five years or less. Panel members take the examination as though they were candidates to appreciate the general level of difficulty of the items and are given an opportunity of making notes regarding any item about which they have concerns or questions. They then are comprehensively trained on the concept of “minimum competency,” are brought into unison (calibrated) on the first 10 test items, and work in sets of 25 to 50 items reviewing and discussing their ratings. Extreme differences in ratings result in the elimination of the question, but the criterion defining “extreme” was not specified in the documentation. In terms of quality assessment of this process, the only thing that can be said here is that appropriate statistical analyses are apparently conducted to determine the degree to which the SMEs were in agreement with each other. The results of these analyses were not provided in the documentation.

The NCCAOM Process

In the materials provided in their March 5, 2004 package, NCCAOM described their methodology. The number of panel members was not specified but the panel was trained to mutually agree on the concept of “minimal competency.” Each exam item is rated independently by each panel member and the ratings are shared with the group. When judgments differ by more than 25%, judges are asked to explain their reasons for the ratings. From the documentation, it appears that while the raters can change their ratings at that time there is no requirement that they do so. Thus, some items could have very different Angoff judgments and remain in the test (unlike the California process that would eliminate such items). The ratings are then compiled in the same manner as was described for California. And, as was the case for California, in terms of quality assessment of this process, the only thing that can be said here is that appropriate statistical analyses are apparently conducted to determine the degree to which the SMEs were consonant with each other. However, the results of these analyses were not provided in the documentation.

Evaluation of the Processes Used by California and NCCAOM

The processes used by both test development groups are not essentially different, and both are pretty standard applications of a basic standard setting criterion-referenced pass point setting process. In that sense, both represent best practice. However, because neither group provided their quality assessment results, it cannot be said how well either group actually accomplished its task.

State Control of the Pass Point Setting Process

By currently using a licensing examination developed and under its own authority, California through OER controls the entire pass point setting process. Continuing to have OER as its internal test development consultant would allow the state to still completely control the process.

If the state decided to adopt the national test, it is reasonable to expect that control of setting the pass point might change to a certain degree. Based on my correspondence with NCCAOM contained in Attachment G, it appears that the passing point NCCAOM establishes is used by all of the states subscribing to its testing program. If a state wished to conduct its own standard setting process, it would require obtaining certain information. This information would include a copy of the exam and scoring key. It might also include the item and test statistics for that exam. Regardless, at least two issues become involved. First, it is not clear that NCCAOM would be willing to provide that information to the state since it is “extremely protective of our examinations and related information. The information is very accessible but only on a need-to-know basis.” It is not clear that a state’s wish to perform its own standard setting process would meet this criterion. Second, if

NCCAOM would agree to supply the necessary information, there would most likely be some cost to the state for being provided with the material.

As an alternative procedure, NCCAOM would apparently be willing to invite state representatives to its cut score analysis meetings. If the state wished to depart from the pass point used by the other states, it appears that this would be state’s decision. The problem is that, without conducting its own standard setting process, the state would, in my opinion, be vulnerable to law suits by candidates (rightfully) claiming that the pass point was not set in a valid manner.

Passing Rates

California Passing Rates

We already know from our quality assessment analysis about the difficulty level of each of the tests. Generally, the difficulty levels for the CALE items concentrate in the .6s, .7s, and .8s. Here, we will present the passing rates for the tests. The California results are documented on the website of the Acupuncture Board. Starting with the January 2002 examination, the data were broken down by those who were first time test takers versus those who had taken (and presumably failed) the test before. Very generally, about two-thirds of the total for the exams were first time test takers and, very generally, their pass rates were approximately twice that of the re-examinees.

A summary of the passing rates is presented in Table 22. Only the overall (all test takers considered) results are presented but they are presented by test date from June 2000 by language group. In the table, both the number of candidates as well as their passing rates are shown.

Table 22
California Passing Rates From
June 2000 Though January 2004
By Language Group

Exam Cycle	English		Mandarin		Korean	
	Number of Examinees	Percent Passing	Number of Examinees	Percent Passing	Number of Examinees	Percent Passing
Jan 2004	272	49%	85	44%	104	60%
Aug 2003	341	66%	142	59%	123	71%
Jan 2003	381	56%	147	56%	92	35%
July 2002	330	50%	180	61%	106	57%
Jan 2002	356	52%	164	51%	123	50%
June 2001	279	51%	154	51%	139	52%
Dec 2000	317	60%	151	54%	114	47%
June 2000	286	57%	148	57%	121	52%
Overall	2562	55.32%	1171	54.77%	922	53.43%

As can be seen from Table 22, from June 2000 through January 2004 the passing rate ranged between 35% and 71%, but most of the rates between 50% and 60%. Over the entire three and one half year period, the rate for the entire data set of 4655 records (which

includes some multiple tests for some candidates) was 54.81%. Interestingly, the overall passing rates for the three language groups was virtually identical although there was a bit more fluctuation in the Korean samples than in the other two.

NCCAOM Passing Rates

We already know from our quality assessment analysis about the difficulty level of each of the tests. The acupuncture modules tend to have difficulty values in the .8s and .9s with the .9 range always more heavily populated. Difficulty levels for the Chinese Herbology modules tend to be concentrated in the .7s, .8s, and .9s but the .8s and .9s are still more populated than the .7 range.

In their March 5, 2004 response, NCCAOM provided information on their passing rates. Because the rates are broken down by module and by language within the Acupuncture and Chinese Herbology modules, it is easiest to display a summary of their data in tabular form, and this is done in Table 23. These results represent the older test structure but should be reflective of what can be expected of the 5-module structure that went into effect in 2004. The full set of results is contained in Attachment H.

The figures in Table 23 should be treated as rough approximations at best. NCCAOM did not supply the number of candidates who sat for each exam. Larger sample sizes would result in more stable percentage values. Those taking the test in Korean test takers have substantially lower passing rates than those taking the test in either English or Chinese.

Table 23
Ranges of NCCAOM Pass Rates From
March 2001 Through October 2003

Exam Module	English		Chinese		Korean	
	% Passing Range	Median	% Passing Range	Median	% Passing Range	Median
Acupuncture	75%–81%	80%	60%–86%	74%	25%–82%	50%
Chinese Herbology	68%–82%	74%	44%–83%	76%	15%–43%	36%

The Point Location module was not broken down by language group and so does not appear in the table. Passing rates for the Point Location module ranged from 74% to 80% with a median pass rate of 78.5%.

Comparison of California and NCCAOM Passing Rates

Without knowing the number of candidates sitting for the national exam, a full comparison between the state and national passing rates cannot be made. However, based on the pass rate data in isolation, it would appear that the passing rate for the NCCAOM exams, with the exception of their Korean test takers, is about 20 percentage points higher than the passing rate for California's license exam. This difference is very roughly in the same direction and at the same order of magnitude that separates the two examination programs in average item difficulty; that is, at a very general level of analysis the NCCAOM modules average about 15 difficulty points higher (easier) than the California exams.

Evaluation of Passing Score Equivalence For California

One very important issue that arises in reviewing passing rates is whether the same level of minimal competence is maintained from one examination cycle to another. If two license exams were administered at two different times, say January and August, the passing scores and/or the passing rates could be the same or different. For example, in 2003 the two California exams had very similar passing scores but somewhat different passing rates (because the mean test score in August was a bit higher than the mean for the January test). The goal of a testing program is to maintain the same competence level as the passing score across their exams.

California supplied us with enough information to determine if the same level of minimal competency was represented in both of their exams. The details of that analysis are presented in Attachment I, but the conclusion may be summarized here. We were able to determine that "...the level of proficiency that was required for passing the January 2003 exam was comparable to what was required for passing the August 2003 exam."

The credit for this outcome falls directly to OER. Within the context of overseeing a license examination program, to demonstrate that the passing points of two or more examination cycles represent the same level of minimal competency is the ultimate goal of such programs. To meet this goal represents the very highest standard of professional testing practice, and OER deserves recognition for this achievement.

We did not receive sufficient information from NCCAOM to engage in a comparable analysis.

The Issue of “Must Pass” Requirements

As the CALE is currently structured, candidates must achieve a particular score on the test to be considered as minimally competent and therefore eligible to receive a license to practice acupuncture. It is therefore possible for candidates to lack knowledge in certain areas, such as the regulations governing the public health and safety, and still pass the exam if they demonstrate considerable knowledge on the other topics. Thus, they may theoretically do quite poorly on, for example, health and safety questions but still be presented with a license to practice by the state.

Other professions in which their practitioners engage in either physical or emotional interaction with their patients appear to have structured their license examination process somewhat differently. These professions seem to have decided that candidates on their license examinations must demonstrate minimal competency in multiple domains rather than just achieving an overall high score before they can acquire a license. The following are just a few examples.

- To become a licensed Physician, medical candidates must separately pass a Step 1 exam covering biology and chemistry, a Step 2 CK exam covering clinical knowledge, and a Step 2 CS exam covering clinical skills. Only then are they eligible to take a Step 3 clinical application exam.
- To become a licensed Dentist, dental candidates must separately pass Part I and Part II of the national boards, then take a clinical examination, and then pass both a written California Dental Law exam and a written test in ethics.
- To become a licensed Psychologist, psychology candidates must first pass the Examination for Professional Practice in Psychology and must then pass the California Jurisprudence and Professional Ethics Examination.
- To become licensed Marriage and Family Therapist, psychology and counseling candidates must first pass a written examination covering a general knowledge of psychology and psychopathology and then pass a written clinical vignette examination.

The practice of acupuncture and oriental medicine involves procedures that directly involve health, safety, and ethical matters. For example, in the course of their normal practice acupuncturists may use implements (e.g., needles) to penetrate the skin of their patients, have their patients ingest certain substances (e.g., herbs), and talk with their patients about topics that are very personal, private, and intimate. Other professions whose practitioners have considerable physical or emotional interaction with their patients or clients build into their licensing process a way to assess for knowledge concerning public safety and ethical issues.

The State of California may decide to either retain the CALE or to join NCCAOM’s national testing program. Regardless of which testing program it ultimately endorses, it is

probably worthwhile to consider the possibility of structuring the acupuncture licensing process in a tiered manner akin to other licensing programs already in effect in California. That is, not only might candidates be required to demonstrate knowledge of the content domains of the discipline of acupuncture and oriental medicine, they might also be asked to demonstrate minimum levels of knowledge regarding public health and safety, as well as ethical issues, before the state is prepared to offer them a license to practice.

Evaluation Summary

This report has explored in detail the stages necessary to develop and administer a high quality license examination program. Within each stage, the portions of the work comprising that stage were discussed and evaluated. Table 24 presents a summary of the evaluations that were made in those discussions.

It may be seen that both examination programs conformed to the standards of professional practice in most instances. It may therefore be more informative to highlight where they differ. Within the occupational analysis, the California process was a bit off the mark, in my opinion, in its statistical calculation of the criticality ratings for tasks and knowledge elements. While a bit irksome, it is not an especially serious problem.

Both the national and state groups did a good job of capturing the task that underlie the professional practice of acupuncture and oriental medicine. For the most part these tasks appear to map well to each other, suggesting that there is at least rough comparability between California and the rest of the nation represented in the national job analysis in what practitioners do in their professional settings. But NCCAOM did not document at all the knowledge elements underlying the tasks. My guess is that these knowledge elements were never documented formally but just formulated in the minds of the item writers as they prepared their test questions. In my opinion, this constitutes a relatively serious breach in the chain of validity evidence for the following reason.

As discussed at length earlier in this report, tasks may remain relatively unchanged over long periods of time but the knowledge base of a discipline steadily grows and changes. Because it is a bit of a “moving target,” it is very important to document what knowledge elements need to be tested by a license examination at any given time. It is this knowledge that is tested in a license examination. Although one can potentially determine that indirectly by studying the actual test questions, the knowledge elements are sufficiently important that they really should, in my opinion, be explicitly and comprehensively articulated in the job analysis. The standards for constructing certification exams argue for this as well. Without the knowledge elements having been documented for the national exam, it is not possible to determine if the two testing programs are actually testing for the same information.

Table 24
Evaluation Summary of the State and
National Examination Programs

Portion of Examination Process		California	NCCAOM
Occupational Analysis	Form expert panels	Consistent with professional standards	Consistent with professional standards
	Identifying tasks	Consistent with professional standards	Consistent with professional standards
	Identify knowledge elements underlying the tasks	Consistent with professional standards	Not explicitly documented
	Organizing the job content	Consistent with professional standards	Consistent with professional standards
	Sampling respondents for job survey	Consistent with professional standards	Consistent with professional standards
	Statistical analysis of important job components	Acceptable but less than best practice	Consistent with professional standards
	Reliability of the job analysis questionnaire		Rater reliability: not computed
		Internal consistency: very high for tasks; very high for knowledge elements	Internal consistency: very high for tasks (no knowledge elements were on the survey)
Developing the exam	Weighting of the content areas	Consistent with professional standards	Consistent with professional standards
	Crafting items	Consistent with professional standards	Not documented
Administering the exam		Consistent with professional standards	Consistent with professional standards
Quality assessment	Test reliability	Consistent with professional standards	Consistent with professional standards
	Item difficulty levels	Moderate	Relatively easy
	Point-biserial correlations	Consistent with professional standards	Consistent with professional standards
	Comparability of language versions	Seems likely	Not documented
Pass point setting	Process used	Consistent with professional standards	Consistent with professional standards
	Passing rate	Around 55%	75%+ except for lower Korean pass rate
	Maintaining same minimal competency standards	Yes	Not documented

California's process of crafting items for the exam was thoroughly documented and represents what is typically done in a high quality testing program. NCCAOM did not document their process at all, but I do not regard this as a problem. Done appropriately, the item writing process is straightforward enough that it need not be documented provided that the product is good; in a sense, the quality of the items speaks for the process that generated them. Given the results of the point-biserial correlations, NCCAOM's item writing process appears to work well.

At the level of analysis that was able to be done on this project, it appears to me that California succeeded in the challenging task of achieving relative comparability of the three language versions of the exam. NCCAOM did not supply the relevant information to determine how well it fared in this endeavor.

Generally, both testing programs produced very reliable exams (not surprising given the length of the tests) using items of good quality as judged by the point-biserial correlations. I would characterize California's test as moderate in difficulty and the national modules as relatively easy.

California tends to pass around 55% of its candidates regardless of language group. NCCAOM tends to pass someplace in the neighborhood of 75% of the candidates who sit for the exam in either English or Chinese, but their Korean samples seem to fare much worse and show a large degree of variability. Since NCCAOM did not supply sample sizes for these subgroups, I can only conjecture that the Korean samples were quite small and thus the poor (or superb) performance of one or two candidates had a dramatic effect on the full sample outcomes. Overall, though, the Korean samples are passing the NCCAOM modules at a much lower rate than the other groups.

Some Consequences of Choosing One Test or the Other

If it is decided that the state will retain the testing program administered by OER not much needs to change (unless it is determined that certain aspects relating to health, safety, and ethics need to be separately tested in a "must pass" or hurdle structure). Further, contained in this report are various recommendations that, in my opinion, can strengthen an already well done testing program if they are implemented.

If it is decided that the state will join the NCCAOM testing program, then several things will follow. The most obvious change will be that California will give up much or virtually all of its control of the testing process. While it may send representatives to participate in the test development process, California will be obliged to use the NCCAOM examination structure. If it wants to supplement those modules with its own health, safety, and ethics exam, it can do that on its own.

The state will also probably wind up using the pass point established nationally. This will likely be the case because it would be a complex and difficult process at best to acquire the test, the key, and the item statistics from NCCAOM for California to conduct its own pass point setting procedures. Without such procedures in place, the Acupuncture Board could not validly establish a different pass point and survive a court challenge to its judgment.

How the content of the NCCAOM examinations would change is rather uncertain. While we know that the tasks and/or content of the profession as practiced in California are very similar to the national tasks and content, it is impossible to say the same regarding the knowledge elements that underlie those tasks. Even when NCCAOM revises its job analysis to include the huge California influx (and the state should not use the national exams until that is done and the tests are revised as necessary), there may still be a difference, however small, between California and the rest of the country. It is possible, to cite one possible example, that the balance between Traditional Chinese Medicine and Non-Traditional Chinese Medicine might be different between California and the rest of the country. In addition, since the only place in the job analysis where any such differences might be detected would be in the knowledge elements, it would be very important that California see the those elements. I therefore strongly suggest that if California does adopt the national examination program, it make as one of its conditions for participating that the new job analysis performed by NCCAOM document the knowledge elements in addition to documenting the tasks and content areas.

The Choice Between the California and National Examination Programs

Based on all of the documentation made available to me for this project, one cannot help but conclude that, despite some weaknesses or documentation failures here and there, both testing programs conscientiously strive toward excellence and have in fact produced very good products. The two testing programs have each captured a weighted composite of the tasks performed in professional practice, have generated items of high quality, and have determined passing criteria in accord with accepted practice.

Nonetheless, the documentation that was provided does allow this writer to distinguish somewhat between these two testing programs. For example, it was possible to determine that a comparable level of minimal competency was maintained between the two California exams administered in 2003. It also appears that there is no substantive difference in either the test statistics or the passing rate of the English, Mandarin, and Korean language groups on the California tests. These are no minor accomplishment and speak extremely well for the quality of OER's testing program. Not enough information was supplied by the national group to perform a similar analysis on their tests.

The identification of the underlying knowledge elements and their linkage to the tasks and to the content of professional practice is a very important component in the evaluation of a license examination and critical in a comparison of two license exams. This sort of documentation provides (a) a critical part of the test documentation to support the validity of the license examination, and (b) the foundation for the development of test questions. Determining the underlying knowledge elements is recognized and—one might judge—even emphasized in the test development standards to which NCCAOM explicitly subscribes; however, these elements were not documented at all in the materials that I reviewed.

While it can be stated with a reasonable degree of confidence that the tasks identified by the national and state occupational analyses are very similar, that is not enough to support the claim that the two examinations are testing for the same knowledge. Since knowledge almost always changes faster than tasks, and since knowledge forms the basis of the tests, it is

necessary to compare the tests on this aspect as well as the tasks. Without that documentation from the national testing program, it is difficult to say that the national exams actually test for the same knowledge elements that are tested in the California examination.

Therefore, I offer my considered opinion based on everything that I thus far have reported: It is my judgment that the California testing program is to be preferred over a very good national testing program. I recommend that OER should be complemented on its high quality work and that this agency be charged with the continued development and administration of the acupuncture license exam. I would also suggest that if some of the small modifications which I have offered to OER in various places in this report are felt to be reasonable, that they be implemented if possible.

Attachment A

Initial Request For Information September 19, 2003

September 19, 2003

Christina S. Herlihy, Ph.D., Chief Executive Officer
National Certification Commission for
Acupuncture and Oriental Medicine
11 Canal Center Plaza, Suite 300
Alexandria, VA 22314

Marilyn Nielson, Executive Officer
Acupuncture and Oriental Medicine Board
1424 Howe Avenue, Suite # 37
Sacramento, CA 95825-3233

With Copies to:

Tracy A. Ferrel, Ph.D., Chief
Office of Examination Resources
501 S Street, Suite 3
Sacramento, CA 95814

Hattie Hanley, MPP, Project Manager
Little Hoover Commission
925 L Street, Suite 805
Sacramento, CA 95814

Stephen Klein, Ph.D.
120 Ocean Park Boulevard # 609
Santa Monica, CA 90405

Dear Dr. Herlihy and Ms. Nielson:

As you probably already know, the Little Hoover Commission of the State of California has been charged by the legislature to answer several questions concerning the licensing of individuals who provide acupuncture and oriental medicine services to the public. One of the legislature's questions relates to the license examination itself, calling for a comparison of the exam used by the state and the national exam. A panel consisting of two experts in test development and validation, Steve Klein and myself, will be performing that analysis for the Little Hoover Commission. Although the paperwork to officially bring us on board is not quite in place, I have been given the green light by the Little Hoover Commission to launch the work at this time.

In preparing for the start this comparison analysis, Steve Klein and I have identified materials and documentation that we need to review. We have therefore compiled a list of items that we are asking the state and the national group to provide to us. As we engage in this work, it is very possible that we will need additional information; thus, we may make one or more additional requests for information from you over the next several months.

We have divided our request into three areas: examination information, test scores of persons who took both exams (double test takers), and an NIH consensus study.

Examination Information

The following list of materials and documentation relates to the development of the exams, outcomes from administrations of the exam, and decisions made on the basis of the exam outcomes. We would appreciate being sent such information for each administration

and version of the test. Because a new national exam format will be introduced in the immediate future, whatever information is currently available for that exam should be included as well in the format listed below. Insofar as it is possible, we are attempting to structure this process so that we have parallel information regarding the state and national exams.

It is also possible that some of the information we are requesting may already be packaged together. We do ask, however, that to the extent that it is possible the materials be divided into separate portions matching the elements in our list below. If there are items you believe are important for us to review but are not included in this initial list, please let us know. Here is this list:

- theory underlying the exam process
- occupational analysis methodology
- methodology for determining critical competencies
- list of critical competencies with accompanying descriptive statistics
- test plan
- linkage of test plan to occupational analysis
- item analyses for last 3 years worth of exams
 - standard/classic item analysis (e.g., item difficulty, distracter analysis, corrected item-total/biserial correlations)
 - stability of item statistics (e.g., item difficulty, item-total correlations) across administrations
 - item characteristic curves (ICCs) based on either theta or total test score
 - group differences (e.g., gender, ethnicity, language) on items
- test analyses for last 3 years worth of exams
 - factor analyses of the test
 - test equating procedures
 - total test characteristic curves (TCCs) based on either theta or total test score
 - IRT analysis
 - comparability of different language forms of the exams
 - group differences (e.g., gender, ethnicity, language) on total test score
 - reliability analyses
- pass-point setting
 - selection of subject matter experts
 - methodology
 - rater agreement statistics

Double Test Takers

We have been informed that there may be several hundred individuals who have taken both the state and the national exams within the past 3 or 5 years. We hope to include an analysis of these data into our work. There are at least a couple of strategies logically available to us to obtain such data.

- Perhaps the most straightforward strategy is for both testing groups to send us a list of test takers with their social security number or other identifier together with each person's score on the exam (or scores on the exam if they have taken the exam more than once). We would match the individuals on the state and national lists and create a data file with that information plus other nonidentifying information. Once built and

verified, we would destroy all identifying information linked to the scores, thereby working with numbers that were no longer traceable back to the individual test takers. Thus, the confidentiality and anonymity of the test takers would be preserved.

•A less desirable strategy is to mail a survey to all test takers in the state who took the California test within our time period. We would then ask them to provide us with their scores on the two exams if they in fact sat for both. The drawbacks to this strategy are many and substantial (e.g., low and selective return rate, inability to authenticate reported scores, risk of individuals being identified by name and address, the expense incurred by the state), and we would prefer to avoid this procedure if possible.

What we need at this juncture is a response from the state and national groups about how best each believes this information can be provided to us. You are welcome to contact me directly on this matter so that we may more efficiently reach a solution to the issue.

NIH Consensus Study

We have also been informed that there was a Consensus Study performed under the auspices of NIH to determine or study the scope of practice of acupuncture and oriental medicine. It would be useful for us to review this document as well. I am hoping that one or more of the recipients of this letter have a copy of it and we would very much appreciate it if this could be provided to Hattie Hanley of the Little Hoover Commission so that they have it for their records. Hattie would then be able to make copies for Steve and me.

Final Word

My intent is to keep the lines of communication open and to communicate with everyone simultaneously on any issues of mutual concern. Most of you know how to reach me but, for the record, my address and phone number are on the letterhead and my e-mail addresses are given below. If you prefer e-mail as the communication medium, please let me know. I do not have electronic addresses for Marilyn Nielson and Tracy Ferrel. My e-mail addresses are:

[e-mail addresses provided]

Please send one copy of your material to Steve at the address above and one copy of your material to me at my home address. Since I am located in the state capitol, we may be able to bypass the U.S. Postal Service or Federal Express by having me acquire my material in person from those in Sacramento—let me know. If you have any questions about any of this, feel free to contact me.

Thank you for your help and cooperation in this matter. I look forward to working with you over the next several months.

Sincerely,

Larry Meyers, Ph.D.

Attachment B

Request to State of California For Further Information January 14, 2004

January 14, 2004

Tracy A. Ferrel, Ph.D., Chief
Office of Examination Resources
501 S Street, Suite 3
Sacramento, CA 95814

With Copies to:

Marilyn Nielson, Executive Officer
Acupuncture and Oriental Medicine Board
Commission
444 N. 3rd Street, Suite 260
Sacramento, CA 95814

Hattie Hanley, Project Manager
Little Hoover
925 L Street, Suite 805
Sacramento, CA 95814

Stephen Klein, Ph.D.
120 Ocean Park Boulevard # 609
Santa Monica, CA 90405

Dear Tracy:

I've started my review of the narrative material you sent to me. This material apparently comprises the documentation, that is, the validity evidence, supporting the California license examination for acupuncture and oriental medicine. I appreciate your having responded to our previous questions and requests for more information regarding some of the statistical analyses on the test items. I am now focusing on the narrative and tabular content of the material.

The document entitled "Acupuncture Practice" was obviously intended to describe the methodology and outcome of the occupational analysis. I have several questions concerning the content of this report as well as some of the other material, and I was hoping that you might be able to clarify a few things for me. Since this material was put together by OER staff, I'm trusting that the information I am asking about but which does not seem to be included in the documentation is still available to you. I'll try to refer you to specific places in the report so that you can see my questions in context.

1. On page 3 (Chapter 2 Survey Questionnaire) under the Interviews heading, we are told that interviews were conducted with some unspecified number of California licensed acupuncturists. There is no indication of the demographics of these individuals or how they were sampled. Since this served as the basis for developing the knowledge and task statements that went on the questionnaire that in turn presumably drove the content of the test, it is important to learn something about these sources. I would like to know how many individuals participated in this process, what their demographics look like, how they were sampled, and any other relevant information concerning them.

2. Still on page 2 under the heading Tasks and Knowledge Statements, the report indicates that the lists of tasks and knowledge statements derived from the initial interviews were reviewed by two separate panels of acupuncturists. Again, no information about these individuals is supplied. I would appreciate

it if you could provide a description of these panels in the same way as I indicated above.

3. The report just indicates in that same paragraph that the task and knowledge statements were placed into appropriate content areas. Now, at this point in the examination planning process, OER is crossing an important threshold in the development of a test plan that turns out to be defined in terms of these categories. It is therefore of great importance for test validation purposes to discuss these categories in detail. Since the report appears to be silent on this matter, I must ask you to explicate the process by which these content areas were determined. Once you have done that, would you please also indicate what evidence exists to support the claim that the statements were properly categorized and how “appropriate placement” was measured.

4. On page 3 under the heading of Questionnaire Distribution, the report indicates that all practitioners who had been licensed for 20 years or less were invited to on-site panels to complete the questionnaire. My questions are: (i) How many invitations were actually sent?; (ii) Why were on-site panels used (with the obvious potential consequences of fewer practitioners showing up who further may not be representative of the profession) rather than just mailing the questionnaire itself to everyone who you deemed appropriate (thus enabling all those who could not or would not leave their practice to provide their input)?

5. On page 5 (Chapter 3 Response Rate and Demographics) under the heading Distribution and Return of Questionnaires, the report indicates that 282 completed questionnaires formed the basis of the evaluation of the task and knowledge statements. The report states “The number of responses that were included in the analysis was sufficient to assume that the results are representative of the profession using a 95% confidence level.” I am puzzled at the wording of this statement. As you know, sheer numbers in the sample do not automatically become representative of the population unless the sample size approaches that of the entire population. Since I am guessing that there were very many more than 300 acupuncturists practicing in the state at the time that this study was conducted, 282 probably does not reflect an awfully large proportion of the population. I would greatly appreciate learning about the evidence and reasoning behind this assertion quoted above.

6. Chapter 3 of the report supplies the demographics of the 282 practitioners who completed the questionnaire. How do these compare with the population of acupuncturists in California if such information is known?

7. On page 11 (Chapter 4 Data Analysis and Results) the task criticality index that was used by OER is discussed. Criticality was apparently computed by multiplying the importance, frequency, and criticality ratings together for each rater and then computing a mean for the index. I have several questions on this:

a. What are the correlations among the rating scales for the task statements? What are the correlations among the rating scales for the knowledge statements? Since there was no discussion of the reasoning concerning the decision to multiply the individual scale ratings together, it would be useful to see how independent these ratings were for each of the two sets.

b. The decision to perform the multiplication for each rater was a bit surprising to me since the individual rating scales contain a zero value. If any of the scales was given such a rating, then the result of the multiplication must be zero as well. This could potentially adversely affect the results. The strategy that I would have expected to be used would have been to generate overall means for each item on each scale first and then, if a criticality index was to be formed by multiplication, to multiply these “overall” values together. I have two questions here: (i) What was the reasoning behind using the strategy that was used and why was it considered better than the alternative strategy?; (ii) how was the issue of zeros used by the raters dealt with?

c. Readers are referred to Appendixes C and D for the average ratings for the task and knowledge statements, respectively. What is shown in these Appendix tables, unfortunately, does not coincide with the way that the criticality index was computed. We are shown the mean for each statement on each rating scale but multiplying them results in a different value than what is shown for the criticality index. There is thus no way to check the validity of the calculations. This is, of course, another reason for computing the criticality index from these overall numbers rather than by individual rater. Since it is not possible for me (or any other reader to verify that the computations were performed correctly, could you either please make such a determination and let me know the result or permit my staff to perform the analysis if the data file is available to us. Could you also supply the standard deviations around each of the means. Although not addressed at all, this statistic does reflect the degree of homogeneity of the raters’ judgments. It can, and probably should, be used as one of the elements in selecting which task and knowledge statements survive into the next stage.

8. On page 8 under the heading Reliability of Ratings, readers are referred to tables showing the reliability coefficients. I’ve got a couple of questions here:

a. Are these Cronbach alpha coefficients? The report is silent on this.

b. In the last paragraph on the page the report states “the reliability coefficients were highly significant ($p < .01$). That is, respondents rated the task and knowledge statements consistently...” Let’s ignore the improper characterization of “highly,” although I do take issue with it. But as you know, being significantly different from zero does not mean that we would judge the ratings to be consistent. When you interpret the magnitude of the coefficients, do you judge them to reflect rater consistency?

9. On page 12 under the heading Differences in Ratings the report states that there were no differences between entry-level (undefined) and practitioners who were licensed for “a longer period of time” (whatever that was), and that there were no differences among the different language groups. Since no analyses or other evidence was supplied here, we have an unsubstantiated assertion. Can you supply evidence to support these assertions?

10. On page 13 (Chapter 5 Description of Practice) the report indicates that the criticality cutoff values were set at 7.33 and 2.85 for task and knowledge statements, respectively.

a. How were these decision points determined?

b. There is a contradiction in the depiction of the cutoff value. Specifically, the report states in the second paragraph "...the criticality index for task statements was 7.33...Thirteen tasks that...had criticality indices less than 7.33...were eliminated." The one-sentence third paragraph states "The task statements with criticality indices below 7.33 are indicated in Appendix C with the phrase *Below Cutoff Value*." This would suggest to me that tasks with criticality indices of 7.33 or higher meet the retention criterion. Yet in Appendix C page 91 we see that Task 11 had a criticality value of exactly 7.33 and was defined (perhaps with others in the same situation) as being below the cutoff value. Please explain how what rule is being used to establish cutoff values here.

11. On page 13 under the heading Linking Task and Knowledge Statements, the assertion is made that some panel of unspecified individuals linked the knowledge statements to their associated task statements.

- a. What was the purpose of this?
- b. What are the results and where are they?
- c. What criterion was required for linkage to be secured?
- d. What became of unlinked statements?

12. Examination development is addressed in the materials only by showing the power point presentation presented by OER on test development. As you know, this really represents an extremely small piece of the validity evidence that is needed to support a license examination. Among the many things we do not know from this report is (a) a detailed description of those individuals who wrote the examination items, and most critically (b) the linkage of the test items back to the task and knowledge statements from the occupational analysis. Please provide this information if it is available.

13. There appears to be no examination plan shown. My research analyst was obliged to consult the Board's website to learn what proportion of the exam was represented by the main content areas since that information is not explicitly dealt with in the report. My third question above asks about the development of these content areas. Assuming that there is validity evidence to support their use in this situation, it is still necessary to derive their relative weights from the occupational analysis. Obviously there are better and worse ways to figure out a weighting scheme. Please indicate how the examination weights were derived in this case.

14. Setting a pass point for the examination is addressed in the materials only by showing the power point presentation presented by OER on that topic. As you know, this really represents an extremely small piece of the validity evidence that is needed to support a license examination. It is clear that OER uses one of the variations of the Angoff procedure for setting pass points. But who participated in the process, the details of the procedure used to facilitate rater agreement, the degree of consistency in the Angoff ratings themselves, how the final Angoff rating per item was set (e.g., mean, trimmed average, median), and the rater consistency for the final Angoff ratings is completely unspecified. If they are available, please provide the details that are missing.

For the moment, these are my questions based on the material that you sent to me. If you need me to clarify any of the above, please feel free to contact me. Thank you in advance for the work you will be doing to respond to my concerns.

Sincerely,

Larry Meyers

Attachment C

**Request For Further Information to NCCAOM
With a Deadline of April 30, 2004 For Receipt
March 28, 2004**

March 28, 2004

Betsy Smith
NCCAOM
11 Canal Center Plaza
Alexandria, VA 22314

cc:

Hattie Hanley, MPP, Project Manager
Little Hoover Commission
925 L Street, Suite 805
Sacramento, CA 95814

Steve Klein
120 Ocean Park Boulevard #609
Santa Monica, CA 90405

Ed O'Neil
UCSF Center for the Health Professions
3333 California Street, Suite 410
San Francisco, CA 94118

Re: Acupuncture and Oriental Medicine exam

Dear Betsy:

Thank you for supplying in your mailing dated March 5, 2004 some of the information that I had requested. I am sorry to pursue my inquiries with you so soon after Christina's death, but there are time constraints involved that are affecting all of us. As you know, I will be providing the Little Hoover Commission with a report in the late spring addressing the acupuncture exams of the State of California and NCCAOM. In order for me to fulfill my obligation, I need to obtain certain materials, information, or documentation from you by the end of April so that I can consider these inputs from you in my report. I will briefly list what I still need from you at this point. If I need any further information, I will let you know as soon as possible.

- In the cover letter from NCCAOM dated March 5, 2004, you stated: "You will note that a number of items in our attached responses are marked as 'not available.' These are items that we do not calculate or analyze, and therefore, we cannot provide this information." I understand that some analyses based on Item Response Theory are not performed and, while they are potentially interesting and useful, I would not define them as critical for my review. Other analyses and information are much more central, and a complete evaluation of the NCCAOM test would not be able to be made without them. Further, I believe that Christina did indicate in a follow-up e-mail that standard or classic item analyses are in fact calculated despite the "not available" response in the letter. **Thus, please send to**

me the standard or classic item analyses you have performed on the exam items in each of your modules for the last two years. These analyses should include, on a module by module basis, item difficulty, corrected item-total correlation, and distracter analyses. I will certainly be willing to accept photocopies of the analyses that have already been completed rather than have you reproduce these analyses from scratch.

- In terms of the maintenance of your item pool, would you please tell me (a) the size of your item pool; (b) the number of new items that are added to the item pool each year; (c) how those new items are pre-tested; (d) approximately how many of the items on each exam (i) have not been used before as scored items, (ii) were used the year before, (iii) were not used for two or three years, and (iv) were not used for four or more years.
- You have described the process you use for test equating. **Please send the results of this procedure from the last couple of years of tests.**
- In the report of your national job analysis (dated August 2003) prepared by Applied Measurement Professionals, Inc. it was stated that 1,265 job analysis surveys were suitable for data analysis and served as the basis for the rest of the test development procedure. Figure 7 on page 13 of the report indicates that 18% of that sample was from the southwest region comprising AZ, CA, HI, and NV. That percentage translates to about 228 respondents. **Please indicate how many of these individuals were from California as opposed to the other three states.**
- Recognizing that the NCCAOM test was intended to be used as a credential or licensing exam at a national level, it is quite reasonable that sampling for the job analysis survey was done in broad geographic regions and that demographic information is presented in the report in aggregated form across these regions. But since we are evaluating the exam for use in California, it should also be understandable that we are particularly interested in the distribution of the demographic variables for our state. **Please provide the breakdown of the demographic information in the same form as shown in the job analysis report for just the California sub-sample of respondents.**
- Considerable detail was spent in the national job analysis report on the obtaining, finalizing, and organizing of the tasks that served as the basis for exam development. But the examination is a set of paper-and-pencil modules, and paper-and-pencil exams ordinarily test for the knowledge, skills, and abilities (KSAs) underlying successful performance of the tasks. That is, your tests do not assess candidates actually performing the tasks (which would be accomplished by some type of “performance” or “hands-on” test) but rather presumably test for the competencies underlying task performance. The material you sent does not appear to describe the process of translating the task statements into the format of a

written credential or licensing test (e.g., identifying the KSAs to be tested, identifying the difficulty level at which the items should be written, and so on).

Please provide the documentation of this very important step in the examination development process.

I must emphasize that I need the information I have requested above so that I can conduct a comprehensive review of your examination process. Since I must submit my report to the Little Hoover Commission by late spring, I am operating under a time constraint the window for which opened in the early fall when I made my original request and which will close very soon. I therefore must ask you to supply me with the requested information by the end of April (i.e., April 30, 2004). If you cannot supply a particular piece of that information, please indicate why so that I may include that reason in my report.

Thank you very much for being responsive to my requests. If you have any questions, please feel free to contact me.

Sincerely,

Larry Meyers

Attachment D

Request For Further Information To State of California March 28, 2004

March 28, 2004

Marilyn Nielsen, Executive Officer
California Acupuncture Board
444 N. 3rd Street, Suite 260
Sacramento, CA 95814

Tracy Ferrel, Ph.D.
Office of Examination Resources
Department of Consumer Affairs
501 S Street, Suite 3
Sacramento, CA 95814

cc:

Hattie Hanley, MPP, Project Manager
Little Hoover Commission
925 L Street, Suite 805
Sacramento, CA 95814

Steve Klein
120 Ocean Park Boulevard #609
Santa Monica, CA 90405

Ed O'Neil
UCSF Center for the Health Professions
3333 California Street, Suite 410
San Francisco, CA 94118

Re: Acupuncture license examination

Dear Marilyn and Tracy:

Thank you for the information you have supplied to me throughout this current acupuncture project. As the time to finish this project draws to a close, I have attempted to determine what other information I need in order to prepare my report to the Little Hoover Commission by a late spring deadline. Thus, having reserved the option of returning to you to ask for additional material, I am now taking the opportunity to do so.

At this time, I am interested in four issues: item pool size and maintenance, the representation of members of the different cultures/languages serving as subject matter experts, the use of two languages in addition to English in the testing process, and the statistical analysis that Tracy's statistical person was to complete. I briefly address each below:

Issue 1

In terms of the maintenance of your item pool, would you please tell me:

- The size of your item pool
- The number of new items that are added to the item pool each year
- Approximately how many of the items on each licensing test (a) have not been used before as scored items, (b) were used the year before, (c) were not used for two or three years, and (d) were not used for four or more years

Issue 2

I understand your position arguing that it is inappropriate to collect information on a variety of demographic variables. You state on page 1 of your March 10 response to me: “However, Government Code section 1870.2 and Civil Service section 1798.14 prohibit Department of Consumer Affairs licensing boards from collecting information that is not necessary to fulfill their regulatory duties (e.g., age, gender, race, and national origin).”

You are clearly attentive to and recognize the importance of these demographic issues when you say on page 2 of your March 10 response to me: “...OER can assure you that the number of men and women included in the occupational analysis interviews was relatively equal. In addition, because this profession is comprised of practitioners from several different cultures, an effort was made to interview practitioners from the major cultures represented (i.e., Chinese, Korean, and English). However, specific information regarding the nationality of individual interviewees was not recorded.”

Given that the the licensing exam is actually administered in three different languages, it is difficult for me to understand how this demographic is considered by OER to be “not necessary” according to Government Code section 1870.2 and Civil Service section 1798.14. It seems to me that the nationality of the individuals you consulted in the process of developing the exam is a very important component of bolstering the validity evidence for the exam. Since you did not record the “heritage” of your interviewees or your subject matter experts (SMEs), perhaps we can make some less precise inferences of their diversity based on last name. Therefore, if this information is available would you please supply me with a list of SMEs who participated in each of the following stages of the testing process:

- The 25 SMEs who were interviewed for developing the survey questionnaire (#1 on page 1 of your March 10 response)
- The two panels who reviewed the task and knowledge statements (#2 on page 2 of your March 10 response)
- Individuals participating in the exam development workshops (#12 on page 12 of your March 10 response)

- Individuals participating in the pass point setting workshops (#14 on page 17 of your March 10 response)

Issue 3

On pages 15–16 of your March 10 response to me, you explain in detail about the process of translating the English language developed licensing exam into Chinese and Korean. This is certainly a challenging task that you clearly take quite seriously. I also acknowledge that you have indicated to me verbally that many candidates are not fully proficient in English.

My interest here is in examining the validity evidence that supports administering the exam in languages other than English. Since this is not a common practice for State licensing examinations, I expected to find information in the occupational analysis that tied such testing in different languages to the job. We know that California is home to a very diverse group of residents but they are obviously served by other professions (e.g., dentists, physicians) whose licensing exams are administered, I believe, only in English. Thus, I would appreciate it if you would please address two aspects of this issue for me:

- Is there a legislative mandate or other legal requirement to administer the license exam in languages other than English? If so, would you please supply the documentation for that mandate (e.g., photocopy of the government code).
- Would you please explicate the validity evidence from the occupational analysis supporting the use of multiple language versions of the acupuncture license exam.

Issue 4

Tracy, I believe that your statistical consultant was going to perform some analyses that I had requested a few months ago and which we discussed when you, I, and Marilyn met. I understand that she was experiencing some health problems, and I do hope that she is faring well, but my recollection is that I have still not received that work. Could you please provide it to me.

I must emphasize that I need the information I have requested above so that I can conduct a comprehensive review of the State's licensing examination process. Since I must submit my report to the Little Hoover Commission by late spring, I am operating under a time constraint the window for which opened in the early fall when I made my original request and which will close very soon. I therefore must ask you to supply me with the requested information by the end of April (i.e., April 30, 2004). If you cannot supply a particular piece of that information, please indicate why so that I may include that reason in my report.

Thank you very much for responding to my requests. If I have additional questions I will be back in touch with you as soon as possible. If you have any questions for me, please feel free to contact me.

Sincerely,

Larry Meyers

Attachment E

**Copy of E-mail From NCCAOM's
Attorney Indicating That The Information
Needed For the Double Test Taker
Analysis Cannot Be Provided
April 16, 2004**

On Friday, April 16, 2004, at 11:57 AM, Jeff.Glassie@shawpittman.com wrote:

Larry:

This follows our telephone call regarding your request to review certain information of NCCAOM in connection with your work for the UCSF and Little Hoover Commission. NCCAOM is willing to provide as much information as it can relevant to the requests in your March 28, 2004 letter, but must have your agreement to and acceptance of the attached Examination Security and Confidentiality Agreement.

In addition, there is certain information and materials that NCCAOM is not permitted to provide to you. Specifically, NCCAOM policy states that "The NCCAOM respects the privacy of all applicant, candidates and Diplomates. All materials submitted or received in connection with application and all test scores are held in confidence, except upon permission for disclosure from the applicant, candidate or Diplomate or except as required by law, including disclosure to governmental licensing bodies upon appropriate written request." Therefore, NCCAOM will not be able to provide individual score reports in response to your request.

Please sign the attached Agreement and return it to me.

Regards, Jeff

Jeff Glassie
Shaw Pittman, LLP
2300 N Street, N.W.
Washington, D.C. 20037
202-663-8036
202-663-8007 (fax)
jeff.glassie@shawpittman.com

(See attached file: T%ZW0!!.DOC)

This electronic message transmission contains information from the law firm Shaw Pittman LLP that may be confidential or privileged. The information is intended solely for the recipient and use by any other party is not authorized. If you are not the intended recipient, be aware that any disclosure, copying, distribution or use of the contents of this information is prohibited. If you have received this electronic transmission in error, please notify us immediately by telephone (1.202.663.8000, or in the UK, 020 7847 9500) or by electronic mail at postmaster@shawpittman.com. Thank You.<T%ZW0!!.DOC>

Attachment F

- 1. E-mail Messages From NCCAOM
Discussing Their Use of
All Items on Their Test
May 18, 2004**

- 2. Excerpt From Materials Dated
April 29, 2004 as a Response
To Request For Further
Information Discussing Item
Scored on NCCAOM's Tests**

1. E- Mail Response

From: Debra Persinger <dpersinger@nccaom.org>
Date: Tue May 18, 2004 9:02:44 AM US/Pacific
To: "Meyers, Lawrence S" <larrym@saclink.csus.edu>, Betsy Smith <bsmith@nccaom.org>
Cc: hattie hanley <hattie.hanley@lhc.ca.gov>, "ed o'neil" <eoneil@itsa.ucsf.edu>, stephen klein <klein@rand.org>
Subject: RE: confidentiality request

Aggregate data is fine to use. Your wonderful explanation (I can tell you are a teacher) has prompted me to wonder whether you can tell from the item analyses if an item was actually used or not. Since we have 20 pretest items per test, if an item had a p value of .99 then it doesn't really do much to discern the competent from the incompetent and likely would not have been scored. Will you take that into account if you have a bunch of items in the .99 region? I suppose the overall test mean would be a better indicator. Maybe I just answered my own question, but I'll send this anyway.

-----Original Message-----

From: Meyers, Lawrence S
Sent: Tuesday, May 18, 2004 11:00 AM
To: Debra Persinger; Betsy Smith
Cc: hattie hanley; ed o'neil; stephen klein
Subject: RE: confidentiality request

Debra

I now know that I intend to include summaries of the statistical information that you sent me on April 29, although I am still working on the formatting of these summaries. I will be aggregating some of the item information, such as the p values and the point-biserials. Most likely, I will indicate how many items fell within a range of values (e.g., how many items had point-biserials between .20 and .29) as well as report the averages across the entire test. I intend to report some of the test statistics as well, such as the overall mean and the reliability of the test. I assure you, however, that no individual item will be identified or even mentioned; in fact, we will not subdivide the tests by content area but will deal only at the full module level. What will be reported will be aggregated statistics only.

According to the confidentiality agreement, I need to check with you to make sure that you are okay with presenting this information in the report. For your information, the corresponding information from the California exams will be included.

Please let me know your judgment on this matter.

Larry

2. Excerpt From April 29, 2004 Response

NCCAOM Item Pool

(a) Size of Item Pool: [response provided by NCCAOM]

(b) Number of New Items Added Each Year: [response provided by NCCAOM]

(c) How Those Items are Pretested: New items are placed on the exam in Active Status. If an item performs unsatisfactorily, the item is not scored. The appropriate examination development committee then reviews the item before it may appear on another examination.

Attachment G

E-mail Messages From NCCAOM Discussing States Joining the National Testing Program May 10–16, 2004

From: Debra Persinger <dpersinger@nccaom.org>
Date: Sun May 16, 2004 6:24:19 AM US/Pacific
To: 'Larry Meyers' <larryerin@comcast.net>, Betsy Smith <bsmith@nccaom.org>
Cc: hattie hanley <hattie.hanley@lhc.ca.gov>, ed o'neil <eoneil@itsa.ucsf.edu>
Subject: RE: another question

There are a number of options depending on what exactly CA would want.

1. There would be NO cost to CA for all the people who have already been through our exam system as a result of having taken the national boards and the CA licensure exam. (It would cost the applicant \$25 to have us send the Exam Result/Certification Status Report.)
2. There would be NO cost to CA for any CA licensure applicant who chooses to apply to us for certification. Again it would cost the applicant \$25 for the score report and the applicant would pay exactly the same fees as any other certification applicant.
3. If CA just wanted to send us a list of names of people that they wanted to test without going through our certification/eligibility review process then this is where a cost would be incurred. Right now the pricing is approximately 1 and 1/2-2X the cost of the exam fee which some states pass on to their applicants. e.g., Point location is \$260 (normally \$130) biomedicine is \$200 (normally \$150). This is for those states who allow folk to be licensed but who cannot meet national standards of eligibility in terms of education or training so they would never be approved by us to test. Those states allow some people to still be licensed as long as they can pass the board exams. It seems unlikely that CA would have anyone in this category and I imagine that all of their applicants could meet our eligibility requirements (based on ACAOM standards) but this is purely supposition on my part. We send the results directly to the state and not to the applicants. There are only a handful of people at most each year that the states send us in this category since the states prefer to just have their applicants go through our system and then apply to the state for licensure once they have met our requirements. It is so much cheaper (free) for the states to have us take care of all the administrative work and eligibility verification, not to mention testing. The states of course then charge their own fees for licensure application. Additionally, the states can rely on our monitoring for disciplinary actions which is info that we share with the states if one of our Diplomates violates standard practice or the Code of Ethics. If CA needed to rely on this category then the volume of candidates would ultimately dictate the pricing - greater volume, less cost per candidate - this would have to be worked out during contract negotiations but I don't see it posing a problem.

Debra A. Persinger, Ph.D.
Interim CEO
NCCAOM phone 703.548.9004 xt 3245 fax 703.548.9079
11 Canal Center Plaza, Suite 300, Alexandria, VA 22314

CONFIDENTIALITY NOTICE: This electronic transmission is intended only for the use of the individual to which it is addressed, and may contain information that is privileged or confidential. If the reader of this message is not the intended recipient, or the employee or agent responsible for delivering the message to the intended recipient, you are hereby notified that any dissemination, distribution or copying of this communication is strictly prohibited. If you have received this communication in error, please notify the sender and delete the message immediately.

-----Original Message-----

From: Larry Meyers [mailto:larryerin@comcast.net]

Sent: Saturday, May 15, 2004 3:00 PM

To: Debra Persinger; Betsy Smith

Cc: hattie hanley; ed o'neil

Subject: another question

Debra and Betsy,

If California wanted to switch over to use your exam process instead of its own for licensing testing, approximately what costs would the state face in joining your program?

Thanks,

Larry

From: Debra Persinger <dpersinger@nccaom.org>

Date: Thu May 13, 2004 10:30:25 AM US/Pacific

To: "Meyers, Lawrence S" <larrym@saclink.csus.edu>, Betsy Smith <bsmith@nccaom.org>

Cc: hattie hanley <hattie.hanley@lhc.ca.gov>, "ed o'neil" <edoneil@itsa.ucsf.edu>, stephen klein <klein@rand.org>

Subject: RE: repeated failure

Hello Larry,

Having pondered this overnight, I still don't think I can come up with a answer, even though you only request an "approximate" cost. All such details and cost analyses would need to be worked out during contract negotiations. My guess would be arbitrary at best so it's probably better that I don't even attempt to make something up.

Rather than turn everything over to an outside party (a gathering of state representatives) my initial instincts are to say that it would more likely remain under our purview wherein we could make all necessary data feely available and the state appointed reps would be invited to be a part of our cut score analysis meetings. You probably get the picture by now that we are extremely protective of our examinations and related information. The information is very accessible but only on a need-to-know basis. This not only serves to protect the security and integrity of the exam process but also to diminish the liability and protect the individuals (including

subject matter experts) who work with the exam.

As always, let me know if I can bring clarity to my stream of consciousness.

Debra A. Persinger, Ph.D.
Interim CEO
NCCAOM phone 703.548.9004 xt 3245 fax 703.548.9079
11 Canal Center Plaza, Suite 300, Alexandria, VA 22314

CONFIDENTIALITY NOTICE: This electronic transmission is intended only for the use of the individual to which it is addressed, and may contain information that is privileged or confidential. If the reader of this message is not the intended recipient, or the employee or agent responsible for delivering the message to the intended recipient, you are hereby notified that any dissemination, distribution or copying of this communication is strictly prohibited. If you have received this communication in error, please notify the sender and delete the message immediately.

-----Original Message-----

From: Meyers, Lawrence S [mailto:larrym@saclink.csus.edu]
Sent: Tuesday, May 11, 2004 12:55 PM
To: Debra Persinger; Betsy Smith
Cc: hattie hanley; ed o'neil; stephen klein
Subject: RE: repeated failure

Debra,

Your response was very helpful. Let me pursue the issue for the moment and pose the hypothetical situation in which a state wished, for whatever reason, to perform their own Angoff procedure. To accomplish that, the state would require the sort of information that I indicated earlier. What would NCCAOM be able to provide and at what very approximate cost? Thanks much for getting back to me on this.

Larry

-----Original Message-----

From: Debra Persinger [mailto:dpersinger@nccaom.org]
Sent: Tue 5/11/2004 6:57 AM
To: Meyers, Lawrence S; Betsy Smith
Cc: hattie hanley; ed o'neil; stephen klein
Subject: RE: repeated failure

I might be able to clarify..... since our exam is competency based, one can either meet the standard (equated cut score of 70) or not meet the standard (by standard I mean the national standard for safe and effective practice). Theoretically, if all candidates meet the standard we could have a 100% pass rate. All of the states that use our exams choose to use the national standard for their licensure purposes. That being said, states,

unlike us, can be the gatekeepers and could choose for example, to have a pass score of 80 instead of our pass score of 70 if they felt that there were too many acupuncturists in their state or some other arbitrary reason of their choosing. That is the prerogative of the states, although none choose this option. Since our job analysis revealed no significant differences in the tasks performed in CA compared to the rest of the nation it probably wouldn't be necessary to replicate the Angoff studies that we perform to set the standard. I hope my 2 cents worth is more of a help than a hindrance.

Debra A. Persinger, Ph.D.
Interim CEO
NCCAOM phone 703.548.9004 xt 3245 fax 703.548.9079
11 Canal Center Plaza, Suite 300, Alexandria, VA 22314

CONFIDENTIALITY NOTICE: This electronic transmission is intended only for the use of the individual to which it is addressed, and may contain information that is privileged or confidential. If the reader of this message is not the intended recipient, or the employee or agent responsible for delivering the message to the intended recipient, you are hereby notified that any dissemination, distribution or copying of this communication is strictly prohibited. If you have received this communication in error, please notify the sender and delete the message immediately.

-----Original Message-----

From: Larry Meyers [mailto:larrym@saclink.csus.edu]
Sent: Monday, May 10, 2004 4:41 PM
To: Betsy Smith
Cc: hattie hanley; ed o'neil; stephen klein; Debra Persinger
Subject: Re: repeated failure

Betsy,

Thanks for this information. Let me drill down a little deeper to make sure that I understand your answer.

For a state to set its own pass point in a valid manner, it would need to perform its own Angoff procedure, it would need a copy of the test, the answer key, and maybe it would also need the item difficulty and item total correlations for each item if they used that information. They would then need to assemble a panel of subject matter experts to engage in the Angoff process under the direction of a testing specialist either from the state or some consultant the state had hired for that purpose. The members of the panel would then examine and potentially discuss each test item to provide their Angoff judgments.

Is it the case that NCCAOM would supply the above information to the state? And would the state be given permission, if indeed it was required, to engage in the process I described above?

On Monday, May 10, 2004, at 12:17 PM, Betsy Smith wrote:

Dear Larry:

Our email went down last Friday and I suspect that is why your emails were returned. Thank you for trying again.

In answer to your question about pass points, yes, we do set the pass point for all of the states. However, this is because they choose to accept our pass point. But, in fact, a state can always elect to set their own passing score. Even though other states accept the scores recommended by NCCAOM's psychometricians, this in no way precludes California from accepting responsibility for setting their own.

I hope this answers your question and that your next email will go through smoothly. Once again, if you have any further questions, please let me know.

With best regards,

Betsy Smith

-----Original Message-----

From: Larry Meyers [mailto:larrym@saclink.csus.edu]

Sent: Monday, May 10, 2004 10:49 AM

To: Betsy Smith

Cc: hattie hanley; ed o'neil; stephen klein

Subject: repeated failure

Betsy,

I have attempted to send this message for the last couple of days from both of my accounts and it has regularly been bumped back to me. I decided to wait until the start of the new week to try again. Here is the text of the message:

>> Thanks. I may have a few questions that you can answer from time to time and will contact you. Here's one: Several states license acupuncturists and obviously use your exam. Does NCCAOM set the pass point for some or all of these states or do you turn over the test and the test results to the state for them to set their own pass point?

Thanks for answering this.

Larry

Attachment H

Passing Percentages Provided by NCCAOM April 29, 2004

NCCAOM 2001 Examination Statistics

Asian Bodywork Therapy	Date	Pass Rate	Total Number of Candidates
English Language	July 2001	82%	56

Acupuncture	Date	Pass Rate	Total Number of Candidates
Chinese Language	March 2001	79%	57
Chinese Language	July 2001	81%	64
Chinese Language	November 2001	74%	76
English Language	March 2001	80%	750
English Language	July 2001	80%	622
English Language	November 2001	75%	722
Korean Language	March 2001	82%	109
Korean Language	July 2001	82%	67
Korean Language	November 2001	52%	50

Point Location	Date	Pass Rate	Total Number of Candidates
All Languages	March 2001	80%	901
All Languages	July 2001	80%	743
All Languages	November 2001	75%	852

Chinese Herbology	Date	Pass Rate	Total Number of Candidates
Chinese Language	March 2001	71%	21
Chinese Language	July 2001	83%	23
Chinese Language	November 2001	77%	17
English Language	March 2001	82%	204
English Language	July 2001	73%	218
English Language	November 2001	74%	258
Korean Language	March 2001	33%	21
Korean Language	July 2001	39%	23
Korean Language	November 2001	15%	27

2002 National Examination Statistics

MARCH 2002		
	# Tested	% Passed
Acupuncture CWE*		
<i>Chinese Language</i>		
Overall	34	74%
<i>English Language</i>		
Overall	631	80%
<i>Korean Language</i>		
Overall	25	32%
Chinese Herbology CWE		
<i>Chinese Language</i>		
Overall	19	79%
<i>English Language</i>		
Overall	222	68%
<i>Korean Language</i>		
Overall	14	29%
Point Location Examination		
<i>All Languages</i>		
Overall	700	74%

JULY 2002**		
	# Tested	% Passed
Acupuncture CWE*		
<i>Chinese Language</i>		
First Time Takers	21	57%
Repeat Takers	4	75%
Overall	25	60%
<i>English Language</i>		
First Time Takers	496	83%
Repeat Takers	72	50%
Overall	568	79%
<i>Korean Language</i>		
First Time Takers	20	75%
Repeat Takers	11	55%
Overall	31	68%
Chinese Herbology CWE		
<i>Chinese Language</i>		
First Time Takers	11	64%
Repeat Takers	2	0%
Overall	13	54%
<i>English Language</i>		
First Time Takers	208	76%
Repeat Takers	27	33%
Overall	235	72%
<i>Korean Language</i>		
First Time Takers	13	46%
Repeat Takers	1	0%
Overall	14	43%
JULY 2002 continued ...		
Point Location Examination		
<i>All Languages</i>		
First Time Takers	511	80%
Repeat Takers	125	69%
Overall	636	77%

NOVEMBER 2002**		
	# Tested	% Passed
Acupuncture CWE		
<i>Chinese Language</i>		
First Time Takers	22	82%
Repeat Takers	4	25%
Overall	26	73%
<i>English Language</i>		
First Time Takers	462	86%
Repeat Takers	65	43%
Overall	527	81%
<i>Korean Language</i>		
First Time Takers	11	55%
Repeat Takers	5	40%
Overall	16	50%
Chinese Herbology CWE		
<i>Chinese Language</i>		
First Time Takers	5	80%
Repeat Takers	1	100%
Overall	6	83%
<i>English Language</i>		
First Time Takers	191	81%
Repeat Takers	37	41%
Overall	228	74%
<i>Korean Language</i>		
First Time Takers	6	17%
Repeat Takers	5	60%
Overall	11	36%
Point Location Examination		
<i>All Languages</i>		
First Time Takers	447	81%
Repeat Takers	76	54%
Overall	523	77%
Asian Bodywork Therapy CWE		
<i>All Languages</i>		
First Time Takers	73	81%
Repeat Takers	5	60%
Overall	78	80%

* Comprehensive Written Examination

** Beginning with the July 2002 administration, NCCAOM separated first time (First Time Takers) and repeat (Repeat Takers) examinees.

2003 National Examination Statistics

Herbology <i>Administration/Language</i> <i>% Passed</i>	Acupuncture		Chinese	
	<i># Tested</i>	<i>% Passed</i>	<i># Tested</i>	
February 2003				
English – first time taker	541	86	135	79
English – repeat taker	75	33	37	54
English – overall	616	79	172	74
Chinese – first time taker	24	79	7	57
Chinese – repeat taker	7	29	2	0
Chinese – overall	31	68	9	44
Korean – first time taker	8	25	3	0
Korean – repeat taker	8	25	7	0
Korean – overall	16	25	10	0
June 2003				
English – first time taker	521	86	232	82
English – repeat taker	115	54	38	26
English – overall	636	80	270	74
Chinese – first time taker	23	91	5	100
Chinese – repeat taker	8	50	3	33
Chinese – overall	31	81	8	75
Korean – first time taker	5	60	2	100
Korean – repeat taker	8	25	5	0
Korean – overall	13	38	7	29
October 2003				
English – first time taker	561	87	193	83
English – repeat taker	109	47	39	26
English – overall	670	80	232	74
Chinese – first time taker	16	94	12	83
Chinese – repeat taker	6	67	3	67
Chinese – overall	22	86	15	80
Korean – first time taker	9	56	1	0
Korean – repeat taker	4	0	6	50
Korean – overall	13	38	7	43

Therapy *Administration Tested % Passed	Point Location		Asian Bodywork	
	# Tested	% Passed	**Administration #	
February 2003			October 2003	
first time taker	526	92	first time taker	61 85
repeat taker	148	36	repeat taker	8 38
overall	674	80	overall	69 80
June 2003				
first time taker	548	82		
repeat taker	115	63		
overall	663	79		
October 2003				
first time taker	586	83		
repeat taker	123	67		
overall	709	80		

*Pass rates for the Point Location Examination are *not* reported by test language.

**The Asian Bodywork Therapy Examination is administered once per year.

Attachment I

Passing Score Equivalence Analysis For California 2003 Exams

PASSING SCORE EQUIVALENCE

In my professional opinion, the level of proficiency that was required for passing the January 2003 exam was comparable to what was required for passing the August 2003 exam. This conclusion is based on the following considerations:

These two exams had almost identical passing scores (116 and 117, respectively) but different passing rates. This difference in passing rates derives from the fact that the mean score in August (across all 175 items) was 5.51 points higher than it was in January. The analyses below examined whether this 5.51 difference stemmed from possible differences in the average difficulty of their questions, differences in the ability of the candidates who took them, or some combination of these two factors.

The mean p-value (i.e., the average proportion of applicants who answered an item correctly) on the 56 items that were common to both exams was .7045 in January and .7361 in August, i.e., a difference of .0316. In other words, when asked the same questions, about 3.2 percent more of the August takers selected the correct choice than did the January takers. This difference is virtually identical to the .0315 difference in mean p-values between these exams on their respective full sets of 175 items. Hence, it may be concluded that the 5.51 difference in overall mean scores between these exams stemmed entirely from the August takers, as a group, being somewhat more proficient than the January takers.

Test Characteristic	January	August	Difference
Passing score	116	117	1.0
Number of takers	623	606	17
Passing rate	53	65	12
Mean p-value			
56 items common to both tests	.7045	.7361	.0316
175 items in full test	.6636	.6951	.0315
Total Test Statistics (175 items)			
Mean score	116.13	121.64	5.51
Standard deviation	17.01	20.14	3.13

The 5.51 difference in average scores between January and August corresponds to about 0.30 standard deviation units ($5.51/18.58 = .30$). In a normal distribution, a passing rate of 53% corresponds to .08 standard deviation units below the mean and a 65% rate corresponds to .38 standard deviation units below the mean. In other words, there is again a 0.30 difference ($.38 - .08 = .30$) between these two tests. In short, the difference in their passing rates coincides with differences in the scores of these two groups on the 56 items that were common to both tests and on their respective full sets of 175 items.

Received from Steve Klein June 18, 2004

Report Supplement

Mapping California To National Tasks From the Occupational Analyses

Tables 1 through 20

**Report Supplement Table I
Content Corresponding to National's
Foundations of Oriental Medicine Module**

National	California
Foundations of Oriental Medicine Module	Patient Assessment
Questioning (collecting relevant information from patient) 1101 Chief Complaint/Current Problem 1102 Medical and Family History 1103 Traditional Questions & Information Gathering	Obtaining Patient History
Assessment (gather data by using the following diagnostic methods – look, listen, smell, touch) 1201 Looking 1202 Listening 1203 Smelling 1204 Touching (palpation)	Performing a Physical Examination Implementing Diagnostic Testing
Analysis & Diagnosis (analyze/classify information collected and establish Oriental medical diagnosis by using traditional Oriental medical theories of physiology and pathology) 1301 Traditional Chinese Medicine 1302 Other Traditions (Japanese, Korean, Worsley)	Developing a Diagnostic Impression Differentiation of Syndromes Forming a Diagnostic Impression
Oriental Medicine Treatment, Planning, Principles, & Strategies 140100 Formulate Treatment Principles 140200 Select Treatment Strategies 1403 Educate Patient 1404 Treat Patient Using Bodywork Techniques	Biomedical Disease Oriental Treatment Planning Forming a Diagnostic Impression
	Providing Acupuncture Treatment Performing Auxiliary Treatment Acupuncture Treatment Contraindications
Professional & Safety Issues 1501 Professional Issues 1502 Safety Issues	Regulations for Public Health and Safety Practice Requirements Patient Protection

**Report Supplement Table 2
Content Corresponding to National's
Acupuncture Module**

National	California
Acupuncture Module	Providing Acupuncture Treatment
Diagnostic Techniques & Treatment & Planning I60100 Ear Diagnosis I602 Appropriate Treatment Theories I603 Points and Sets of Points I604 Microsystem Points	Point Selection Principles Point Categories in Acupuncture Treatment Implementing Microsystems in Treatment
Treatment I701 Identify Treatment Techniques or Modes of Administration I702 Treat Patient I703 Apply Moxibustion I704 Apply Heat/Cold I70500 Electroacupuncture I706 Manage Emergency Situations I70700 Ensure Clean Needle Technique	Point Location and Needling Technique Performing Auxiliary Treatment Acupuncture Treatment Contraindications Treatment Observation and Modification <hr/> Developing a Diagnostic Impression Biomedical Disease <hr/> Regulations for Public Health and Safety Patient Protection
Treatment Evaluation I80100 Evaluate the Results of Treatment I80200 Make Appropriate Modifications and Recommendations I80300 Consult with Patient on Additional Areas of Concern	Providing Acupuncture Treatment Treatment Observation and Modification Performing Auxiliary Treatment

**Report Supplement Table 3
Content Corresponding to National's
Chinese Herbology Module**

National	California
Chinese Herbology Module	Prescribing Herbal Medicinals
Planning & Treatment (select or formulate treatment strategies based on diagnosis or evaluation) 2201 Identification 2202 Selection 2203 Formulation	Identification of Herbs Prescribing and Administering Herbs
Implementation 230100 Gauge Patient's Receptivity and Anticipate Patient's Reaction to Treatment 230200 Assemble the Herbal Formula 2303 Distribute the Herbal Formula(s) and Consult with Patient 2304 Apply Moxibustion 230500 Apply External Herbal Applications (compresses, liniments, plasters) 230600 Treat Patient Using Nutritional Supplements	Prescribing and Administering Herbs
	Providing Acupuncture Treatment
	Performing Auxiliary Treatment
Safety 240100 Identify Precautions & Contraindications 240200 Consider Herb-Drug Interactions 240300 Consider Herb-Herb Interactions 240400 Consider Herb-Food Interactions and the Use of Herb in Food	Prescribing Herbal Medicinals Identification of Herbs Prescribing and Administering Herbs
Treatment Evaluation 250100 Evaluate the Results of Treatment 250200 Make Appropriate Modifications and Recommendations 250300 Consult with Patient on Additional Areas of Concern	Prescribing Herbal Medicinals
	Prescribing and Administering Herbs

**Report Supplement Table 4
Content Corresponding to National's
Biomedicine Module**

National	California
Biomedicine Module	Patient Assessment
Questioning 1901 Chief Complaint/Current Problem 1902 Medical and Family History 1903 Traditional Questions and Information Gathering	Obtaining Patient History Evaluating for Western Pharmacology
Western Medical Assessment: consider the results of the following diagnostic evaluations as well as knowledge of anatomy, physiology, and pathology, to help assess the patient's health status, communicate effectively with other members of the health care team, and make referrals as indicated. 2001 Western Medical Assessment	Implementing Diagnostic Testing Developing a Diagnostic Impression
	Biomedical Disease
Legal, Professional, & Safety Issues 2101 Legal Issues: Follow local/state/federal laws/rules, regulations, and statutes 2102 Professional Issues 2103 Safety Issues	Biomedical Disease Regulations for Public Health and Safety
	Practice Requirements Patient Protection

**Report Supplement Table 5
Content Corresponding to National's
Point Location Module**

National	California
Point Location Module	Providing Acupuncture Treatment
"All acupoints may be accessed."	Point Selection Principles Point Categories in Acupuncture Treatment Implementing Microsystems in Treatment Point Location and Needling Technique Acupuncture Treatment Contraindications

Note: The National did not provide explicit test specifications for this module. Thus there is some uncertainty as to which California subareas match those that are part of National's Point Location module. We have included the larger subareas rather than the task statements as California matches.

Report Supplement Table 6.a
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Questioning with
California's Patient Assessment – Obtaining Patient History

National	California
I101 Chief Complaint/Current Problem	
I10101 Identifying information (e.g., name, address, phone, age, gender, marital status, occupation)	
I10102 Chief/secondary complaints (e.g., frequency, severity, symptoms, onset, duration, functional changes)	T1. Assess patient's presenting complaints by obtaining info re symptoms to determine focus of examination.
I10103 History, assessment, and treatment of current condition	T2. Assess general medical status by interviewing patient re health history to determine effect on presenting complaint.
I10104 Patient's treatment goals	

Report Supplement Table 6.b
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Questioning with
California's Patient Assessment – Obtaining Patient History

National	California
I102 Medical and Family History	
<ul style="list-style-type: none"> • I10201 Medical history 	T2. Assess general medical status by interviewing patient re health history to determine effect on presenting complaint.
<ul style="list-style-type: none"> • I10202 Family medical history 	T2. Assess general medical status by interviewing patient re health history to determine effect on presenting complaint.
<ul style="list-style-type: none"> • I10203 Medications • I10204 Neurological symptoms 	
<ul style="list-style-type: none"> • I10205 Social and personal relationships and life goals 	T3. Identify impact of emotional factors by evaluating significant events in patient's life to determine contribution to symptom development.
<ul style="list-style-type: none"> • I10206 Domestic violence, including sexual abuse 	T3. Identify impact of emotional factors by evaluating significant events in patient's life to determine contribution to symptom development.
<ul style="list-style-type: none"> • I10207 Patient boundaries and concerns about physical contact 	
<ul style="list-style-type: none"> • I10208 Memory and cognitive processes 	
<ul style="list-style-type: none"> • I10209 Factors that interfere with data gathering (e.g., drugs, food, drink) 	

Report Supplement Table 6.c
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Questioning with
California’s Patient Assessment – Obtaining Patient History

National	California
I 103 Traditional Questions and Information Gathering	
<ul style="list-style-type: none"> • I 10301 Sleep patterns 	T4. Identify sleep patterns to determine the cause and effect on pattern development.
<ul style="list-style-type: none"> • I 10302 Breathing and respirations 	
<ul style="list-style-type: none"> • I 10303 Skin 	
<ul style="list-style-type: none"> • I 10304 Bowel movements, digestion, appetite, and thirst 	T8. Evaluate digestion by identifying gastrointestinal responses to determine Middle Jiao function. T9. Evaluate level of appetite by determining patient’s preferences for food to determine nature of condition. T13. Evaluate thirst by determining patient’s preferences for liquids to determine nature of condition. T17. Evaluate bowel function to determine nature of imbalance.
<ul style="list-style-type: none"> • I 10305 Secretions and excretions 	T16. Evaluate urine characteristics to determine nature of imbalance. T17. Evaluate bowel function to determine nature of imbalance.
<ul style="list-style-type: none"> • I 10306 Bleeding and bruising 	
<ul style="list-style-type: none"> • I 10307 Tastes (bitter, sour, etc.) 	T12. Evaluate preferences for or aversions to flavors or temperatures to determine nature of imbalance.
<ul style="list-style-type: none"> • I 10308 Nutritional levels and patterns (e.g., quantity of food consumed, food supplementation, regularity of eating) 	T6. Gather info re diet by evaluating nutritional habits to determine contribution to pattern development. T10. Identify eating patterns to determine impact on digestive functioning.
<ul style="list-style-type: none"> • I 10309 Temperature 	T18. Evaluate for the presence of fever or chills to determine nature of disharmony.
<ul style="list-style-type: none"> • I 10310 Perspiration/sweating 	T19. Evaluate patterns of perspiration to determine nature of disharmony.
<ul style="list-style-type: none"> • I 10311 Pain 	T22. Evaluate nature of pain to determine etiology and pathology.
<ul style="list-style-type: none"> • I 10312 Emotions 	T3. Identify impact of emotional factors by evaluating significant events in patient’s life to determine contribution to symptom development.

Report Supplement Table 6.c (continued)

National	California
I 103 Traditional Questions and Information Gathering	
<ul style="list-style-type: none"> • I 10313 Stressors 	T3. Identify impact of emotional factors by evaluating significant events in patient's life to determine contribution to symptom development.
<ul style="list-style-type: none"> • I 10314 Use of alcohol, tobacco, caffeine, and non-pres. drugs 	T5. Gather info re environmental influences by asking questions re exposures to determine impact on pattern development. T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I 10315 Exercise and physical activity 	T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I 10316 Sexual activity 	T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I 10317 Libido 	T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I 10318 Birth-control methods 	T14. Evaluate gynecological history to determine imbalances.
<ul style="list-style-type: none"> • I 10319 Menstrual, gynecologic and obstetric history 	T14. Evaluate gynecological history to determine imbalances. T15. Evaluate urogenital history to determine imbalances.
<ul style="list-style-type: none"> • I 10320 Male reproductive health 	T15. Evaluate urogenital history to determine imbalances.
<ul style="list-style-type: none"> • I 10321 Dizziness and tinnitus 	
<ul style="list-style-type: none"> • I 10322 Palpitations or chest constriction 	
<ul style="list-style-type: none"> • I 10323 Vision, hearing, and speech 	T20. Evaluate eye function by asking questions regarding ocular changes. T21. Evaluate auditory functioning by asking questions regarding changes in acuity.
<ul style="list-style-type: none"> • I 10324 Edema 	

Report Supplement Table 7.a
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Assessment with
California’s Patient Assessment – Performing a Physical Examination

National	California
I201 Looking	
• I20101 General physical appearance	
• I20102 Face	T24. Observe the face and eyes to determine the outward manifestation of the Shen. T25. Observe face for distinguishing characteristics to aid in pattern discrimination.
• I20103 Eyes	T24. Observe the face and eyes to determine the outward manifestation of the Shen.
• I20104 Tongue	T36. Evaluate tongue coating by examining qualities to identify abnormalities in functioning. T37. Analyze variations in tongue body to determine systemic disharmonies.
• I20105 Skin and complexion	T31. Identify condition of the cutaneous region by examining skin to determine evidence of obstruction or injury. T32. Identify dermatological condition by examining skin to determine evidence of obstruction or injury.
• I20106 Nails and hands	T31. Identify condition of the cutaneous region by examining skin to determine evidence of obstruction or injury.
• I20107 Hair	
• I20108 Ears	
• I20109 Spirit/shen (expression and general behavior)	T23. Evaluate level of energy by observing patient’s demeanor to determine quality of Qi. T24. Observe the face and eyes to determine the outward manifestation of the Shen.
• I20110 Body structure (e.g., constitution, weight, structural imbalance, individual body tissues)	T26. Evaluate constitution by observing physical characteristics to determine Five Element associations.
• I20111 Posture (e.g., center of gravity, imbalance between left and right, front and back)	T35. Perform orthopedic assessment by evaluating neuromuscular skeletal systems to identify pathology.
• I20112 Movement (e.g., gait, fluidity, range of motion)	T33. Perform range of motion examination to identify areas of restricted movement.
• I20113 Symptom site/local area of complaint	
• I20114 Secretions and excretions	T29. Assess phlegm characteristics to identify nature of pathogenic influence.
• I20115 Factors that interfere with data gathering (e.g., make-up, perfume, lighting)	

Report Supplement Table 7.b
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Assessment with
California's Patient Assessment – Performing a Physical Examination and
Implementing Diagnostic Testing

National	California
I202 Listening	
<ul style="list-style-type: none"> I20201 Sound/tone of voice, including volume and tonal qualities 	T27. Evaluate voice by listening for tonal qualities and strength to determine nature of disharmony.
<ul style="list-style-type: none"> I20202 Abdominal sounds 	T51. Perform auscultation to identify cardiopulmonary or abdominal pathologies.
<ul style="list-style-type: none"> I20203 Respiratory sounds 	T30. Evaluate pulmonary efficiency by assessing respiration to differentiate between patterns.
<ul style="list-style-type: none"> I20204 Manner, pattern, theme, and content of speech 	
<ul style="list-style-type: none"> I20205 Vomiting sounds 	
<ul style="list-style-type: none"> I20206 Hiccups, belching 	
<ul style="list-style-type: none"> I20207 Joint sounds 	T53. Perform neurological examination by evaluating reflexes and cutaneous sensation to identify pathology.

Report Supplement Table 7.c
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Assessment with
California's Patient Assessment – Performing a Physical Examination

National	California
I203 Smelling	
<ul style="list-style-type: none"> I20301 General odor of body 	
<ul style="list-style-type: none"> I20302 Secretions and discharges 	
<ul style="list-style-type: none"> I20303 Mouth/breath odor 	T36. Evaluate tongue coating by examining qualities to identify abnormalities in functioning. T37. Analyze variations in tongue body to determine systemic disharmonies.
<ul style="list-style-type: none"> I20304 Excretions 	T29. Assess phlegm characteristics to identify nature of pathogenic influence.

Report Supplement Table 7.d
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Assessment with
California's Patient Assessment – Performing a Physical Examination and
Implementing Diagnostic Testing

National	California
I204 Touching (palpations)	
• I20401 Abdomen	T52. Perform abdominal palpation to identify organ pathology.
• I20402 Qualities and positions of radial pulse	T38. Assess radial pulse qualities by palpation to determine disharmony.
• I20403 Comparison of regional pulse sites (e.g., carotid, radial, umbilical)	T38. Assess radial pulse qualities by palpation to determine disharmony.
• I20404 Channels and points (including Ashi points)	
• I20405 Shu/Mu Points	
• I20406 Changes in temperature, moisture, texture, sensitivity, tissue structure	
• I20407 Ears	
• I20408 Changes in nature of pain and numbness, with palpation	T39. Palpate areas of tenderness to determine the quality and nature of patient's pain.
• I20409 Passive range of motion	T33. Perform range of motion examination to identify areas of restricted movement. T34. Palpate joints to assess functional integrity.
• I20410 Nodules and tumors	

Report Supplement Table 8.a
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Analysis and Diagnosis with
California’s Developing a Diagnostic Impression – Differentiation of Syndromes and
Forming a Diagnostic Impression

National	California
I301 Traditional Chinese Medicine	
<ul style="list-style-type: none"> • I30101 Five Phase/Element Theory 	T65. Determine Five Element imbalances by evaluating systems of correspondences to determine impact on patient condition.
<ul style="list-style-type: none"> • I30102 Yin Yang Theory 	T56. Identify principle patterns by evaluating symptoms to determine balance of Yin and Yang. T67. Determine Eight Principles diagnosis by evaluating symptom parameters to determine the nature and strength of disease.
<ul style="list-style-type: none"> • I30103 Channel Theory 	
<ul style="list-style-type: none"> • I30104 Eight Extraordinary Channels Theory 	
<ul style="list-style-type: none"> • I30105 Pathological Point findings (e.g., Kokatsu (Kori), Ashi) 	
<ul style="list-style-type: none"> • I30106 Eight Parameters 	T67. Determine Eight Principles diagnosis by evaluating symptom parameters to determine the nature and strength of disease.
<ul style="list-style-type: none"> • I30107 Zang Fu Theory 	T66. Determine Zang Fu diagnosis by evaluating patterns in symptomatology to identify affected Organs.
<ul style="list-style-type: none"> • I30108 Essential substances (Blood, Qi, Fluid, Spirit, Essence) 	T68. Determine relative strength of Qi and Blood by evaluating diagnostic findings to determine the nature of the condition.
<ul style="list-style-type: none"> • I30109 Internal, external and miscellaneous causes of diseases 	T57. Determine causative factors by integrating diagnostic info to identify source of underlying disease patterns.
<ul style="list-style-type: none"> • I30110 Six Stages 	T70. Determine depth of penetration of pathogen by using Six Stage differentiation to develop a treatment strategy.
<ul style="list-style-type: none"> • I30111 Four Divisions/Levels (Wei, Qi, Ying, Xue) 	T69. Determine depth of penetration of pathogen by using Four Level differentiation to develop a treatment strategy.
<ul style="list-style-type: none"> • I30112 San Jiao (Triple Warmer) 	T71. Determine location of pathogenic factor by evaluating patterns of disease manifestation of the San Jiao.
<ul style="list-style-type: none"> • I30113 Differentiation of disease (Bian Bing) 	
	T72. Determine Jing Ye diagnosis by integrating diagnostic findings.

Report Supplement Table 8.b
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Analysis and Diagnosis with
California’s Developing a Diagnostic Impression – Differentiation of Syndromes and
Forming a Diagnostic Impression

National	California
I302 Other Traditions (e.g., Japanese, Korean, Worsley)	
<ul style="list-style-type: none"> I30201 Five Phase/Element Theory 	T65. Determine Five Element imbalances by evaluating systems of correspondences to determine impact on patient condition.
<ul style="list-style-type: none"> I30202 Yin Yang Theory 	T56. Identify principle patterns by evaluating symptoms to determine balance of Yin and Yang. T67. Determine Eight Principles diagnosis by evaluating symptom parameters to determine the nature and strength of disease.
<ul style="list-style-type: none"> I30203 Channel Theory 	
<ul style="list-style-type: none"> I30204 Eight Extraordinary Channels Theory 	
<ul style="list-style-type: none"> I30205 Pathological Point findings (e.g., Kokatsu, Ashi) 	
<ul style="list-style-type: none"> I30206 Twelve Officials Theory 	
<ul style="list-style-type: none"> I30207 Zang Fu Theory 	T66. Determine Zang Fu diagnosis by evaluating patterns in symptomatology to identify affected Organs.
<ul style="list-style-type: none"> I30208 Eight Parameters 	T67. Determine Eight Principles diagnosis by evaluating symptom parameters to determine the nature and strength of disease.
<ul style="list-style-type: none"> I30209 Essential substances (Blood, Qi, Fluid, Spirit, Essence) 	T68. Determine relative strength of Qi and Blood by evaluating diagnostic findings to determine the nature of the condition.
<ul style="list-style-type: none"> I302110 Internal, external, and miscellaneous causes of diseases 	T57. Determine causative factors by integrating diagnostic info to identify source of underlying disease patterns.
<ul style="list-style-type: none"> I302111 Six Stages 	T70. Determine depth of penetration of pathogen by using Six Stage differentiation to develop a treatment strategy.
<ul style="list-style-type: none"> I302112 Four Divisions/Levels (Wei, Qi, Ying, Xue) 	T69. Determine depth of penetration of pathogen by using Four Level differentiation to develop a treatment strategy.
<ul style="list-style-type: none"> I302113 San Jiao (Triple Warmer) 	T71. Determine location of pathogenic factor by evaluating patterns of disease manifestation of the San Jiao.
<ul style="list-style-type: none"> I302114 Sho/Confirmation 	
<ul style="list-style-type: none"> I302115 Constitutional Theory 	
	T72. Determine Jing Ye diagnosis by integrating diagnostic findings.

Report Supplement Table 9.a
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Oriental Medicine Treatment,
Planning, Principles, and Strategies with California's Developing a Diagnostic
Impression – Oriental Treatment Planning and Forming a Diagnostic Impression

National	California
I40100 Formulate Treatment Principles	T81. Determine treatment principle by evaluating patterns of disharmony to develop treatment objectives.
	T55. Identify interrelationship of affected Organs by evaluating constitutional weaknesses or changes resulting from pathogenic influences.
	T57. Determine causative factors by integrating diagnostic information to identify source of underlying disease patterns.
	T58. Integrate symptoms of physiological systems to determine stage of disease progression.
	T59. Identify severity of condition by evaluating level and movement of pathogenic penetration.
	T60. Identify affected channel by evaluating diagnostic info to determine disharmony.
	T61. Differentiate between root and branch of disease by evaluating symptoms to determine focus of treatment.
	T64. Integrate diagnostic findings to form clinical impressions to describe patient's current health status.

Report Supplement Table 9.b
Mapping California to National Tasks from the Occupational Analyses:
National's Foundation of Oriental Medicine Module – Oriental Medicine Treatment,
Planning, Principles, and Strategies with California's Developing a Diagnostic
Impression – Oriental Treatment Planning and Forming a Diagnostic Impression

National	California
I40200 Select Treatment Strategies	T82. Develop a treatment plan by formulating a plan of action to address therapeutic needs of the patient.
	T62. Differentiate between primary and secondary conditions by prioritizing symptoms to develop treatment strategy.
	T63. Determine acute pernicious influences by evaluating for symptoms of illness to determine treatment strategy.

Report Supplement Table 9.c
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Oriental Medicine Treatment, Planning, Principles, and Strategies with California’s Developing a Diagnostic Impression – Biomedical Disease and Providing Acupuncture Treatment – Performing Auxiliary Treatment

National	California
I403 Educate Patient	
<ul style="list-style-type: none"> • I40301 Basic Oriental medicine dietary principles 	T124. Recommend dietary changes by identifying specific foods to add/omit to support treatment.
<ul style="list-style-type: none"> • I40302 Basic nutritional principles 	T124. Recommend dietary changes by identifying specific foods to add/omit to support treatment.
<ul style="list-style-type: none"> • I40303 Pre-treatment orientation 	
<ul style="list-style-type: none"> • I40304 Follow-up instructions 	
<ul style="list-style-type: none"> • I40305 Basic instruction/training references 	
<ul style="list-style-type: none"> • I40306 Basic stretching, movement, and exercise principles and techniques 	T122. Recommend adjunctive therapies patient can implement to support treatment.
<ul style="list-style-type: none"> • I40307 Basic breathing and relaxation principles and techniques 	T122. Recommend adjunctive therapies patient can implement to support treatment.
<ul style="list-style-type: none"> • I40308 Self-treatment techniques 	T123. Recommend lifestyle changes patient can implement to restore or maintain health.
<ul style="list-style-type: none"> • I40309 Lifestyle implications and considerations 	T123. Recommend lifestyle changes patient can implement to restore or maintain health.
<ul style="list-style-type: none"> • I40310 Body mechanics 	T73. Provide patient with information regarding physiological systems to explain how the body functions.
<ul style="list-style-type: none"> • I40311 Ergonomics 	
<ul style="list-style-type: none"> • I40312 Meditation 	T122. Recommend adjunctive therapies patient can implement to support treatment.
<ul style="list-style-type: none"> • I40313 Qi Gong (i.e., explain benefits) 	
<ul style="list-style-type: none"> • I40314 Integration of Oriental medical theory and modern lifestyles 	T74. Inform patient of Oriental medical diagnosis by relating Oriental concepts to Western medicine concepts.
<ul style="list-style-type: none"> • I40315 The healing process 	
<ul style="list-style-type: none"> • I40316 Evaluating change 	

Report Supplement Table 9.d
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Oriental Medicine Treatment, Planning, Principles, and Strategies with California’s Providing Acupuncture Treatment – Performing Auxiliary Treatment and Acupuncture Treatment Contraindications

National	California
I404 Treat Patient Using Bodywork Techniques	
<ul style="list-style-type: none"> • I40401 Acupressure 	T121. Perform Oriental soft tissue techniques by manipulating affected areas to activate functional changes. T134. Identify conditions contraindicated for Oriental soft tissue techniques to avoid deleterious treatment effects.
<ul style="list-style-type: none"> • I40402 Shiatsu 	T121. Perform Oriental soft tissue techniques by manipulating affected areas to activate functional changes. T134. Identify conditions contraindicated for Oriental soft tissue techniques to avoid deleterious treatment effects.
<ul style="list-style-type: none"> • I40403 Tuina 	T121. Perform Oriental soft tissue techniques by manipulating affected areas to activate functional changes. T134. Identify conditions contraindicated for Oriental soft tissue techniques to avoid deleterious treatment effects.

Report Supplement Table 10.a
Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Professional and Safety Issues
with California’s Regulations for Public Health and Safety – Practice Requirements
and Patient Protection

National	California
1501 Professional Issues	
<ul style="list-style-type: none"> • 150101 Follow ethical standards (e.g., NCCAOM Code of Ethics) 	<p>T153. Advertise services by adhering to legal guidelines for disseminating info re treatment provided.</p> <p>T154. Implement therapeutic techniques congruent with professional competence to provide services within scope of practice.</p> <p>T156. Report known or suspected abuse by contacting protective services to comply with mandated reporting requirements.</p>
<ul style="list-style-type: none"> • 150102 Maintain appropriate record-keeping practices 	<p>T152. Maintain patient records by recording treatments given and progress made to track therapeutic progress.</p> <p>T159. Respond to legal mandates for treatment info by releasing patient records to comply with court order.</p>
<ul style="list-style-type: none"> • 150103 Observe established practice management procedures 	
<ul style="list-style-type: none"> • 150104 Maintain confidentiality 	<p>T155. Maintain client’s confidentiality by securing treatment records to protect client’s right to privacy.</p>
<ul style="list-style-type: none"> • 150105 Obtain informed consent 	<p>T157. Obtain informed consent by providing information re treatment benefits, risks, and side effects.</p> <p>T158. Obtain patient’s written consent to disclose treatment information to protect right to privacy.</p>
<ul style="list-style-type: none"> • 150106 Maintain certifications (e.g., NCCAOM, CPR) 	

Report Supplement Table 10.b

**Mapping California to National Tasks from the Occupational Analyses:
National’s Foundation of Oriental Medicine Module – Professional and Safety Issues
with California’s Regulations for Public Health and Safety – Practice Requirements
and Patient Protection**

National	California
I502 Safety Issues	
<ul style="list-style-type: none"> • I50201 Ensure equipment maintenance and safety 	T160. Implement clean needle technique by following guidelines to prevent the spread of pathogens. T162. Dispose of needles and supplies by placing in required container to prevent accidental punctures. T163. Dispose of contaminated material containers by adhering to OSHA requirements for disposal.
<ul style="list-style-type: none"> • I50202 Identify and implement infection control precautions (e.g., universal precautions). 	T161. Implement universal precautions by following guidelines during treatment to prevent cross contamination.

Report Supplement Table 11.a

**Mapping California to National Tasks from the Occupational Analyses:
National’s Acupuncture Module – Diagnostic Techniques and Treatment
and Planning with California’s Providing Acupuncture Treatment – Implementing
Microsystems in Treatment**

National	California
I60100 Ear Diagnosis	T126. Select auricular points by identifying clinical indications to treat patient condition.

Report Supplement Table 11.b
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Diagnostic Techniques and Treatment
and Planning with California's Providing Acupuncture Treatment – Point
Categories in Acupuncture Treatment

National	California
I602 Appropriate Treatment Theories	
<ul style="list-style-type: none"> • I60201 Tonification, supplementation, reinforcement, dispersion / sedation / drainage 	
<ul style="list-style-type: none"> • I60202 Yin Yang Theory 	
<ul style="list-style-type: none"> • I60203 Channels (Meridians) 	T98. Select Lower He-Sea (Uniting) points to connect channel with respective Fu Organs. T103. Select Intersecting/Crossing points to treat diseases manifesting in multiple channels.
<ul style="list-style-type: none"> • I60204 Sheng (generation) and Ke (control) Cycles 	
<ul style="list-style-type: none"> • I60205 Four Needle Concept 	T111. Select Mother/Son (Four Needle Technique) points by identifying complimentary points to address imbalances.
<ul style="list-style-type: none"> • I60206 Mu and Shu 	T95. Select Front-Mu (Alarm) points to address systemic imbalances of an acute nature. T96. Select Back-Shu (Associated) points to address systemic imbalances of a chronic nature. T97. Select coupled points by combining Front-Mu (Alarm) and Back-Shu (Associated) points to balance Yin and Yang.
<ul style="list-style-type: none"> • I60207 Causative Factor 	
<ul style="list-style-type: none"> • I60208 Eight Extraordinary Channel Theory 	T101. Select Confluent points of the Eight Extra Channels by evaluating symptoms to treat conditions of the miscellaneous channels.
<ul style="list-style-type: none"> • I60209 Six Division Theory 	
	T99. Combine points from different categories to provide balanced treatment. T108. Select the Eight Influential Points to affect related anatomical areas (tissues). T110. Select Four Seas points to affect corresponding Qi, Blood, Nourishment, and Marrow.

Report Supplement Table 11.c
Mapping California to National Tasks from the Occupational Analyses:
National’s Acupuncture Module – Diagnostic Techniques and Treatment
and Planning with California’s Providing Acupuncture Treatment – Point
Categories in Acupuncture Treatment, Implementing Microsystems
in Treatment, and Point Selection Principles

National	California
I603 Points and Sets of Points	
<ul style="list-style-type: none"> • I60301 Five Phase/Element Points (control points) 	T100. Select Five Shu (Five-Transporting) points to treat imbalances of the Five Elements.
<ul style="list-style-type: none"> • I60302 Antique Points (Jing Well, Ying, Spring, etc.) 	
<ul style="list-style-type: none"> • I60303 Mu (Alarm) and Shu (Associated) Points 	T95. Select Front-Mu (Alarm) points to address systemic imbalances of an acute nature. T96. Select Back-Shu (Associated) points to address systemic imbalances of a chronic nature. T97. Select coupled points by combining Front-Mu (Alarm) and Back-Shu (Associated) points to balance Yin and Yang.
<ul style="list-style-type: none"> • I60304 Confluent Points of the Eight Extra Channels (Meridians) 	T101. Select Confluent points of the Eight Extra Channels by evaluating symptoms to treat conditions of the miscellaneous channels.
<ul style="list-style-type: none"> • I60305 Sheng (generation) and Ke (control) Cycles 	
<ul style="list-style-type: none"> • I60306 Meeting Points 	
<ul style="list-style-type: none"> • I60307 Entry and Exit Points 	
<ul style="list-style-type: none"> • I60308 Window of the Sky Points 	
<ul style="list-style-type: none"> • I60309 Four Needle Technique 	T111. Select Mother/Son (Four Needle Technique) points by identifying complimentary points to address imbalances.
<ul style="list-style-type: none"> • I60310 Xi-Cleft Points 	T107. Select Xi-Cleft (Accumulation) points to treat acute conditions of the related channel and corresponding Organs.
<ul style="list-style-type: none"> • I60311 Yuan (Source) and Luo (Connecting) Points 	T104. Select Luo-Connecting points that access Divergent channels to strengthen internally-externally related channels. T105. Select Luo-Connecting points to treat conditions associated with paired Yin and Yang channels. T106. Select Yuan-Source (Primary) points to access fundamental Qi for the channel.
<ul style="list-style-type: none"> • I60312 Aggressive Energy Treatment 	

Report Supplement Table 11.c (continued)

National	California
<ul style="list-style-type: none"> • I60313 Extra Points 	T102. Select Extra points to treat corresponding conditions.
<ul style="list-style-type: none"> • I60314 Trigger or Motor Points 	
<ul style="list-style-type: none"> • I60315 Ashi Points 	
<ul style="list-style-type: none"> • I60316 Dong's (Tong's), Tan's (12X12) Points, or Naso/Muno Points/Regions 	
	<p>T83. Develop a point prescription based on treatment principles to correct imbalances.</p> <p>T84. Select distal/proximal points to address affected channels.</p> <p>T85. Select local points by evaluating clinical indication to treat condition.</p> <p>T86. Select adjacent point near the affected area to augment treatment protocol.</p> <p>T87. Select points from different channels to combine treatment of root and branch.</p> <p>T88. Select contralateral points to balance treatment prescription.</p> <p>T89. Select points above and below to balance treatment prescription.</p> <p>T90. Select front and back points to enhance treatment prescription.</p> <p>T91. Select points in the center to treat conditions occurring in the extremities.</p> <p>T92. Select points on the extremities to treat conditions occurring in the center.</p> <p>T93. Select points along the muscle channels by identifying correspondences between Ashi points and features of the muscle channel system.</p> <p>T94. Select points along corresponding channels to affect Cutaneous regions.</p>

Report Supplement Table 11.d

Mapping California to National Tasks from the Occupational Analyses: National's Acupuncture Module – Diagnostic Techniques and Treatment and Planning with California's Providing Acupuncture Treatment – Implementing Microsystems in Treatment

National	California
I604 Microsystem Points	
<ul style="list-style-type: none"> • I60401 Ear points (Nogier, Chinese, NADA) 	T126. Select auricular points by identifying clinical indications to treat patient condition.
<ul style="list-style-type: none"> • I60402 Scalp points (Chinese and YNSA) 	T125. Select scalp points by identifying clinical indications to treat patient condition.
<ul style="list-style-type: none"> • I60403 Foot points 	
<ul style="list-style-type: none"> • I60404 Wrist/ankle points 	

Report Supplement Table 12.a

**Mapping California to National Tasks from the Occupational Analyses:
National’s Acupuncture Module – Treatment with California’s Providing
Acupuncture Treatment – Point Location and Needling Technique
and Acupuncture Treatment Contraindications**

National	California
1701 Identify Treatment Techniques or Modes of Administration	
<ul style="list-style-type: none"> • 170101 Identify correct point location (e.g., anatomical, proportional) 	TI12. Locate points by implementing anatomical landmarks and proportional measurements to identify area for needle insertion.
<ul style="list-style-type: none"> • 170102 Positioning of patient 	TI14. Position patient by moving into recommended configuration to provide for proper needle insertion.
<ul style="list-style-type: none"> • 170103 Consider anatomy to determine depth, precautions, and contraindications. 	TI15. Insert needle according to standard depths to accurately stimulate point.
<ul style="list-style-type: none"> • 170104 Consider precautions related to treatment (e.g., intradermal needles, moxibustion, electricity, guasha, bleeding). 	TI30. Identify conditions contraindicated for needling by evaluating condition and constitution to avoid injury. TI31. Identify conditions contraindicated for electroacupuncture to determine alternative treatment strategy. TI32. Identify conditions contraindicated for cupping to avoid deleterious treatment effects. TI33. Identify conditions contraindicated for moxibustion by evaluating condition and constitution to avoid injury. TI35. Identify conditions contraindicated for adjunctive therapies to determine alternative treatment strategy. TI36. Identify conditions contraindicated for dietary therapy to avoid deleterious treatment effects.

Report Supplement Table 12.b
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment with California's Providing
Acupuncture Treatment – Point Location and Needling Technique and
Performing Auxiliary Treatment

National	California
1702 Treat Patient	
<ul style="list-style-type: none"> • 170201 Acupuncture (inserting needles) 	T113. Identify length of needle retention by evaluating patient constitution and condition to determine time needed for efficacy of result. T116. Apply therapeutic needle technique by manipulating needle to produce intended effect.
<ul style="list-style-type: none"> • 170202 Cups 	T119. Perform cupping technique by placing instrument over area to increase effectiveness of treatment.
<ul style="list-style-type: none"> • 170203 Ear balls/seeds/pellets 	
<ul style="list-style-type: none"> • 170204 Nutritional supplements 	T124. Recommend dietary changes by identifying specific foods to add/omit to support treatment.

Report Supplement Table 12.c
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment with California's Providing
Acupuncture Treatment – Performing Auxiliary Treatment and Acupuncture
Treatment Contraindications

National	California
1703 Apply Moxibustion	
<ul style="list-style-type: none"> • 170301 Direct moxa (e.g., Chinese, non-scarring) 	T117. Apply moxibustion to enhance the effects of treatment. T133. Identify conditions contraindicated for moxibustion by evaluating condition and constitution to avoid injury.
<ul style="list-style-type: none"> • 170302 Indirect moxa (e.g., stick or pole moxa) 	T117. Apply moxibustion to enhance the effects of treatment. T133. Identify conditions contraindicated for moxibustion by evaluating condition and constitution to avoid injury.

Report Supplement Table 12.d
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment

National	California
1704 Apply Heat/Cold	
• 170401 Heating lamps	
• 170402 Hydrocollator packs	
• 170403 Microwave heat pads	
• 170404 Chemical heat pads	
• 170405 Herbal heating pads	
• 170406 Spray and stretch (vapor coolant)	
• 170407 Ice packs	
• 170408 Hot compresses	
• 170409 Cold compresses	
• 170410 Other methods	

Report Supplement Table 12.e
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment with California's Providing
Acupuncture Treatment – Performing Auxiliary Treatment and Acupuncture
Treatment Contraindications

National	California
170500 Electroacupuncture	T118. Perform electroacupuncture by electrically stimulating selected points to enhance effectiveness of treatment. T131. Identify conditions contraindicated for electroacupuncture to determine alternate treatment strategy.

Report Supplement Table 12.f
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment with California's Developing
a Diagnostic Impression – Biomedical Disease and Providing Acupuncture
Treatment – Treatment Observation and Modification

National	California
1706 Manage Emergency Situations	T78. Identify life-threatening conditions by evaluating signs and symptoms to refer to emergency medical treatment T128. Monitor patient's reactions to treatment by evaluating changes in vital signs to identify indications of distress.
• 170601 Fainting	
• 170602 Broken needle	
• 170603 Stuck needle	
• 170604 Organ puncture (e.g., pneumothorax)	
• 170605 Burns	
• 170606 Bleeding	
• 170607 Cardiac or respiratory arrest	
• 170608 Other situations	

Report Supplement Table 12.g
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment with California's Regulations
for Public Health and Safety – Patient Protection

National	California
170700 Ensure Clean Needle Technique	T160. Implement clean needle technique by following guidelines to prevent the spread of pathogens.

Report Supplement Table 13.a
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment Evaluation with California's
Providing Acupuncture Treatment – Treatment Observation and Modification

National	California
180100 Evaluate the Results of Treatment	T128. Monitor patient's reactions to treatment by evaluating changes in vital signs to identify indications of distress.

Report Supplement Table 13.b
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment Evaluation with California's
Providing Acupuncture Treatment – Treatment Observation and Modification

National	California
180200 Make Appropriate Modifications and Recommendations	T129. Reevaluate patient condition by examining changes in symptomatology to determine adjustments to treatment plan.

Report Supplement Table 13.c
Mapping California to National Tasks from the Occupational Analyses:
National's Acupuncture Module – Treatment Evaluation with California's
Providing Acupuncture Treatment – Performing Auxiliary Treatment

National	California
180300 Consult with Patient on Additional Areas of Concern	T122. Recommend adjunctive therapies patient can implement to support treatment.

Report Supplement Table 14.a
Mapping California to National Tasks from the Occupational Analyses:
National’s Chinese Herbology Module – Planning and Treatment with
California’s Prescribing Herbal Medicinals – Identification of Herbs and
Prescribing and Administering Herbs

National	California
2201 Identification	
<ul style="list-style-type: none"> • 220101 Recognize crude herbs 	T139. Determine herbal composition of formulas by identifying chief and deputy herbs to provide hierarchy of ingredients.
<ul style="list-style-type: none"> • 220102 Recognize products that are manufactured with GMP standards 	
<ul style="list-style-type: none"> • 220103 Identify individual herb preparations (e.g., dry fried, honey fried, calcined) 	
<ul style="list-style-type: none"> • 220104 Understand properties and dosage of individual herbs (i.e., nature, taste, directionality, channels entered and actions, toxicity, effects of preparation, and dosage) 	T137. Identify characteristics of herbs and formulas by evaluating attributes to determine therapeutic effect. T144. Identify dosage of herbal prescriptions by evaluating condition and constitution.
<ul style="list-style-type: none"> • 220105 Identify principles governing the combination of herbs (e.g., 18 contraindications and 19 incompatibilities) 	T138. Distinguish between herbs and formulas from same classification by identifying unique features to differentiate effectiveness for condition. T140. Identify complementary herb qualities and point functions to provide integrated treatment. T142. Identify contraindications for herbs and formulas by assessing a patient’s constitution and situation.

Report Supplement Table 14.b
Mapping California to National Tasks from the Occupational Analyses:
National’s Chinese Herbology Module – Planning and Treatment with
California’s Prescribing Herbal Medicinals – Identification of Herbs and
Prescribing and Administering Herbs

National	California
2202 Selection	
<ul style="list-style-type: none"> • 220201 Appropriate herbal formula 	T143. Prescribe herbs and formulas by matching clinical indications with patient condition to treat pathology.
<ul style="list-style-type: none"> • 220202 Differentiate characteristics of Western vs. Chinese herbs 	T141. Identify similarities between herbal prescriptions and Western medications by evaluating therapeutic properties to augment patient treatment.
<ul style="list-style-type: none"> • 220203 Internal and/or external formulas 	
<ul style="list-style-type: none"> • 220204 Formula preparation (e.g., raw herbs, powder, pills, Tinctures, liniment) 	

Report Supplement Table 14.c
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Planning and Treatment with
California's Prescribing Herbal Medicinals – Identification of Herbs and
Prescribing and Administering Herbs

National	California
2203 Formulation	
<ul style="list-style-type: none"> • 220301 Internal structures and dynamics of herbal formulas 	T139. Determine herbal composition of formulas by identifying chief and deputy herbs to provide hierarchy of ingredients. T138. Distinguish between herbs and formulas from same classification by identifying unique features to differentiate effectiveness for condition.
<ul style="list-style-type: none"> • 220302 Indications and functions of representative herb formulas 	T137. Identify characteristics of herbs and formulas by evaluating attributes to determine therapeutic effect. T140. Identify complementary herb qualities and point functions to provide integrated treatment. T142. Identify contraindications for herbs and formulas by assessing a patient's constitution and situation.
<ul style="list-style-type: none"> • 220303 Adjust formula to fit individual cases 	T149. Modify herbal prescription by evaluating changes in patient symptoms to address changing conditions. T150. Modify herbal prescription by identifying secondary or underlying symptoms to address additional conditions.

Report Supplement Table 15.a
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation with California's
Prescribing Herbal Medicinals – Prescribing and Administering Herbs

National	California
230100 Gauge Patient's Receptivity and Anticipate Patient's Reaction to Treatment	T145. Evaluate patient response by assessing for changes to determine herbal treatment modifications.

Report Supplement Table 15.b
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation with California's
Prescribing Herbal Medicinals – Prescribing and Administering Herbs

National	California
230200 Assemble the Herbal Formula	T147. Assemble herbal formulas by combining proportions of effective dosages of herbs. T148. Combine herbs and formulas to address complex patterns.

Report Supplement Table 15.c
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation with California's
Prescribing Herbal Medicinals – Prescribing and Administering Herbs

National	California
2303 Distribute Herbal Formula(s) and Consult with Patient	
<ul style="list-style-type: none"> • 230301 Treatment strategies and goals for herbal treatments 	T143. Prescribe herbs and formulas by matching clinical indications with patient condition to treat pathology.
<ul style="list-style-type: none"> • 230302 Herbal preparation 	T151. Advise patient on herbal preparations by providing instructions to produce intended therapeutic effect.
<ul style="list-style-type: none"> • 230303 Dosage and frequency 	T144. Identify dosage of herbal prescriptions by evaluating condition and constitution.
<ul style="list-style-type: none"> • 230304 Duration of treatment 	
<ul style="list-style-type: none"> • 230305 Side effects and interactions 	T146. Monitor effects of herbs when combined with Western medications to determine potential side effects.

Report Supplement Table 15.d
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation with California's
Providing Acupuncture Treatment – Performing Auxiliary Treatment

National	California
2304 Apply Moxibustion	
<ul style="list-style-type: none"> • 230401 Direct moxa (e.g., Chinese, non-scarring) 	T117. Apply moxibustion to enhance the effects of treatment.
<ul style="list-style-type: none"> • 230402 Indirect moxa (e.g., stick or pole moxa) 	

Report Supplement Table 15.e
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation

National	California
230500 Apply External Herbal Applications (e.g., compresses, liniments, plasters)	

Report Supplement Table 15.f
Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Implementation

National	California
230600 Treat Patient Using Nutritional Supplements	

Report Supplement Table 16.a

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Safety with California's Prescribing
Herbal Medicinals – Identification of Herbs**

National	California
240100 Identify Precautions / Contraindications	T142. Identify contraindications for herbs and formulas by assessing a patient's constitution and situation.

Report Supplement Table 16.b

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Safety with California's Prescribing
Herbal Medicinals – Prescribing and Administering Herbs**

National	California
240200 Consider Herb-Drug Interactions	T146. Monitor effects of herbs when combined with Western medications to determine potential side effects.

Report Supplement Table 16.c

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Safety with California's Prescribing Herbal
Medicinals – Identification of Herbs and Prescribing and Administering Herbs**

National	California
240300 Consider Herb-Herb Interactions	T142. Identify contraindications for herbs and formulas by assessing a patient's constitution and situation. T148. Combine herbs and formulas to address complex patterns.

Report Supplement Table 16.d

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Safety with California's Prescribing
Herbal Medicinals – Identification of Herbs**

National	California
240400 Consider Herb-Food Interactions and the Use of Herb in Food	T142. Identify contraindications for herbs and formulas by assessing a patient's constitution and situation.

Report Supplement Table 17.a

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Treatment Evaluation with
California's Prescribing Herbal Medicinals – Prescribing and Administering Herbs**

National	California
250100 Evaluate the Results of Treatment	T145. Evaluate patient response by assessing for changes to determine herbal treatment modifications.

Report Supplement Table 17.b

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Treatment Evaluation with
California's Prescribing Herbal Medicinals – Prescribing and Administering Herbs**

National	California
250200 Make Appropriate Modifications and Recommendations	T145. Evaluate patient response by assessing for changes to determine herbal treatment modifications. T149. Modify herbal prescription by evaluating changes in patient symptoms to address changing condition. T150. Modify herbal prescription by identifying secondary or underlying symptoms to address additional conditions.

Report Supplement Table 17.c

**Mapping California to National Tasks from the Occupational Analyses:
National's Chinese Herbology Module – Treatment Evaluation with
California's Prescribing Herbal Medicinals – Prescribing and Administering Herbs**

National	California
250300 Consult with Patient on Additional Areas of Concern	T145. Evaluate patient response by assessing for changes to determine herbal treatment modifications.

Report Supplement Table 18.a

**Mapping California to National Tasks from the Occupational Analyses:
National's Biomedicine Module – Questioning with California's Patient
Assessment – Obtaining Patient History**

National	California
1901 Chief Complaint/Current Problem	
<ul style="list-style-type: none"> • 190101 Current biomedical diagnosis 	T1. Assess patient's presenting complaints by obtaining information regarding symptoms to determine focus of examination.
<ul style="list-style-type: none"> • 190102 History, assessment, and treatment of current condition 	T2. Assess general medical status by interviewing patient regarding health history to determine effect on presenting complaint.

Report Supplement Table 18.b
Mapping California to National Tasks from the Occupational Analyses:
National's Biomedicine Module – Questioning with California's Patient
Assessment – Obtaining Patient History and Evaluating for Western Pharmacology

National	California
1902 Medical and Family History	
<ul style="list-style-type: none"> • 190201 Medical history (e.g., previous diagnosis, illnesses, surgeries, hospitalizations, allergies, accidents) 	T2. Assess general medical status by interviewing patient regarding health history to determine effect on presenting complaint.
<ul style="list-style-type: none"> • 190202 Family medical history 	T2. Assess general medical status by interviewing patient regarding health history to determine effect on presenting complaint.
<ul style="list-style-type: none"> • 190203 Medications (e.g., prescriptions, over-the-counter, herbs, dose, term of use) 	T44. Identify types of Western medications patient is taking to determine impact of pharmaceuticals on condition. T45. Identify actions of Western pharmacological agents to determine systems involved. T46. Identify effects and side effects of Western medications to determine need to refer for reevaluation.
<ul style="list-style-type: none"> • 190204 Neurological symptoms (e.g., neuropathies) 	

Report Supplement Table 18.c
Mapping California to National Tasks from the Occupational Analyses:
National's Biomedicine Module – Questioning with California's
Patient Assessment – Obtaining Patient History

National	California
1903 Traditional Questions and Information Gathering	
<ul style="list-style-type: none"> • 190301 Sleep patterns 	T4. Identify sleep patterns to determine the cause and effect on pattern development.
<ul style="list-style-type: none"> • 190302 Breathing and respirations 	
<ul style="list-style-type: none"> • 190303 Skin 	
<ul style="list-style-type: none"> • 190304 Bowel movements, digestion, appetite, and thirst 	T9. Evaluate level of appetite by determining patient's preferences for food to determine nature of condition. T13. Evaluate thirst by determining patient's preferences for liquids to determine nature of condition. T17. Evaluate bowel function to determine nature of imbalance.
<ul style="list-style-type: none"> • 190305 Secretions and excretions (e.g., vomit, sputum) 	T16. Evaluate urine characteristics to determine nature of imbalance. T17. Evaluate bowel function to determine nature of imbalance.
<ul style="list-style-type: none"> • 190306 Bleeding and bruising 	
<ul style="list-style-type: none"> • 190307 Tastes (bitter, sour, etc.) 	T12. Evaluate preferences for or aversions to flavors or temperatures to determine nature of imbalance.
<ul style="list-style-type: none"> • 190308 Nutritional levels and patterns 	T6. Gather info re diet by evaluating nutritional habits to determine contribution to pattern development. T10. Identify eating patterns to determine impact on digestive functioning.
<ul style="list-style-type: none"> • 190309 Temperature (e.g., sensations of hot or cold) 	T18. Evaluate for the presence of fever or chills to determine nature of disharmony.
<ul style="list-style-type: none"> • 190310 Perspiration/sweating 	T19. Evaluate patterns of perspiration to determine nature of disharmony.
<ul style="list-style-type: none"> • 190311 Pain 	T22. Evaluate nature of pain to determine etiology and pathology.
<ul style="list-style-type: none"> • 190312 Emotions 	T3. Identify impact of emotional factors by evaluating significant events in patient's life to determine contribution to symptom development.

Report Supplement Table 18.c (continued)

National	California
<ul style="list-style-type: none"> • I90313 Stressors 	T3. Identify impact of emotional factors by evaluating significant events in patient's life to determine contribution to symptom development. T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I90314 Use of alcohol, tobacco, caffeine, and non-pres. drugs 	T5. Gather info re environmental influences by asking questions re exposures to determine impact on pattern development. T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I90315 Exercise and physical activity 	T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I90316 Sexual activity 	T7. Gather info re lifestyle to determine contribution to symptom development.
<ul style="list-style-type: none"> • I90317 Libido 	
<ul style="list-style-type: none"> • I90318 Birth-control methods 	T14. Evaluate gynecological history to determine imbalances.
<ul style="list-style-type: none"> • I90319 Menstrual, gynecologic and obstetric history 	T14. Evaluate gynecological history to determine imbalances. T15. Evaluate urogenital history to determine imbalances.
<ul style="list-style-type: none"> • I90320 Male reproductive health 	T15. Evaluate urogenital history to determine imbalances.
<ul style="list-style-type: none"> • I90321 Dizziness and tinnitus 	
<ul style="list-style-type: none"> • I90322 Palpitations or chest constriction 	
<ul style="list-style-type: none"> • I90323 Vision, hearing, and speech 	T20. Evaluate eye function by asking questions regarding ocular changes. T21. Evaluate auditory functioning by asking questions regarding changes in acuity.
<ul style="list-style-type: none"> • I90324 Edema 	

Report Supplement Table 19.a
Mapping California to National Tasks from the Occupational Analyses:
National’s Biomedicine Module – Western Medical Assessment with
California’s Patient Assessment – Implementing Diagnostic Testing and
Developing a Diagnostic Impression – Biomedical Disease

National	California
2001 Western Medical Assessment	
<ul style="list-style-type: none"> • 200101 Consider results of Western physical examinations 	T50. Measure vital signs to identify baseline values and pathologies. T51. Perform auscultation to identify cardiopulmonary or abdominal pathologies. T52. Perform abdominal palpation to identify organ pathology. T75. Evaluate symptoms to determine indications of Western conditions that require referral for treatment. T78. Identify life-threatening conditions by evaluating signs and symptoms to refer to emergency medical treatment.
<ul style="list-style-type: none"> • 200102 Ask patient about self examinations (e.g., breast, testicular) 	
<ul style="list-style-type: none"> • 200103 Measure range of motion 	
<ul style="list-style-type: none"> • 200104 Consider results of orthopedic and neurological tests 	T53. Perform neurological examination by evaluating reflexes and cutaneous sensation to identify pathology.
<ul style="list-style-type: none"> • 200105 Consider results of laboratory tests (e.g., blood, stool, urine, PAP smear) 	T47. Evaluate results of laboratory panels by reviewing ranges of values.
<ul style="list-style-type: none"> • 200106 Consider results of imaging tests (e.g., MRI, x-ray, CT scan, colonoscopy) 	T48. Evaluate results of radiographic imaging tests by reading report to identify suspected pathology.
	T76. Prepare reports re patient condition by translating Oriental medical diagnosis into terminology common to other health care providers.
	T77. Interact with health care providers to integrate treatment.

Report Supplement Table 20.a

**Mapping California to National Tasks from the Occupational Analyses:
National’s Biomedicine Module – Legal, Professional, and Safety Issues
with California’s Regulations for Public Health and Safety – Practice
Requirements and Patient Protection**

National	California
2101 Legal Issues	
<ul style="list-style-type: none"> • 210101 OSHA 	T163. Dispose of contaminated material containers by adhering to OSHA requirements for disposal.
<ul style="list-style-type: none"> • 210102 Reporting requirements (e.g., abuse, disease transmission) 	T156. Report known or suspected abuse by contacting protective services to comply with mandated reporting requirements. T159. Respond to legal mandates for treatment info by releasing patient records to comply with court order.
<ul style="list-style-type: none"> • 210103 Biohazard management 	T160. Implement clean needle technique by following guidelines to prevent the spread of pathogens. T161. Implement universal precautions by following guidelines during treatment to prevent cross contamination. T162. Dispose of needles and supplies by placing in required container to prevent accidental punctures. T163. Dispose of contaminated material containers by adhering to OSHA requirements for disposal.
<ul style="list-style-type: none"> • 210104 Fire 	
<ul style="list-style-type: none"> • 210105 Licensure 	T154. Implement therapeutic techniques congruent with professional competence to provide services within scope of practice.

Report Supplement Table 20.b

**Mapping California to National Tasks from the Occupational Analyses:
National’s Biomedicine Module – Legal, Professional, and Safety Issues with
California’s Regulations for Public Health and Safety – Patient Protection**

National	California
2102 Professional Issues	
<ul style="list-style-type: none"> • 210201 Maintain professional hygiene • 210202 Use ICD codes/Insurance 	
<ul style="list-style-type: none"> • 210203 Educate patient regarding appropriate referral to medical and social institutions. 	T157. Obtain informed consent by providing information regarding treatment benefits, risks, and side effects.

Report Supplement Table 20.c
Mapping California to National Tasks from the Occupational Analyses:
National's Biomedicine Module – Legal, Professional, and Safety Issues with
California's Developing a Diagnostic Impression - Biomedical Disease

National	California
2103 Safety (Identify, Manage, and Take Appropriate Action)	T78. Identify life-threatening conditions by evaluating signs and symptoms to refer for emergency medical treatment.
• 210301 Fainting	
• 210302 Burns	
• 210303 Bleeding	
• 210304 Cardiac or respiratory arrest (performing CPR)	
• 210305 Anaphylaxis	
• 210306 Other situations	