Osteopathic Manipulative Treatment Technique Scores on the COMLEX-USA Level 2-PE: An Analysis of the Skills Assessed

Larissa L. Smith, PhD; Wei Xu, MA; Jeanne M. Sandella, DO; and Dennis J. Dowling, DO

Context: The National Board of Osteopathic Medical Examiners uses the Osteopathic Manipulative Treatment (OMT) Global Rating Tool to score candidate performance in OMT during standardized patient encounters. To the authors’ knowledge, no validity research has been published on the underlying constructs measured by this instrument to date.

Objective: To evaluate whether the OMT Global Rating Tool measures 1 holistic skill or multiple skill sets.

Methods: Structural equation models were fit to data from first-time test-takers who took the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE) between July 7, 2014, and April 30, 2015. Two theoretical models were evaluated: a model that posited a single underlying skill set, and a model that posited 2 related but distinct skill sets.

Results: The 1-factor and 2-factor models were fit to data from 4673 third- and fourth-year osteopathic medical students. Fit statistics indicated that the data were best fit by a model representing 2 factors: pre- and post–OMT technique assessment and OMT technique ($r=0.47$).

Conclusion: The OMT Global Rating Tool seems to measure 2 distinct but related skills. Although the results of this study are applicable to broad populations and cannot be used to provide skill-specific subscores, it might be helpful to remediation efforts to target pre- and post–OMT technique assessments and OMT techniques as distinct areas for intervention.

doi:10.7556/jaoa.2016.080
Scoring and assessment instruments in high-stakes evaluation should possess reasonable and empirically demonstrated levels of reliability and validity. For such empirical demonstrations, an understanding of the underlying statistical structure of the instrument is of critical importance. An assessment tool that measures many different skills (eg, communicating with patients, communicating with colleagues, working in a team) may be poorly served by statistical techniques that assume that an assessment tool measures one thing and measures it with highly intercorrelated items.

The National Board of Osteopathic Medical Examiners (NBOME) uses the Osteopathic Manipulative Treatment (OMT) Global Rating Tool in the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE) to assess the OMT skills of candidates. Although the OMT Global Rating Tool is intended to lead to a single global OMT score for any given encounter, and while evidence for the utility of that global score exists, internal analyses by the NBOME have indicated that the tool may be measuring 2 distinct skill sets: pre- and post-OMT technique assessment and the performance of the OMT technique itself.

We used structural equation modeling to compare models with 1 and 2 underlying factors and evaluated which model fit the data better. Structural equation modeling allows researchers to test a set of hypothesized relationships among variables against the empirical relationships when measurement variance in the observed variables is accounted for. In the case of the OMT Global Rating Tool, structural equation modeling can test the hypothesis that there are 2 separate skill sets underlying scores on the tool’s 6 OMT assessment components, and that those skill sets are related but cannot be considered equivalent.

This study seeks to obtain a better understanding of the measurement model underlying the indicators of OMT performance used in the OMT Global Rating Tool. Although structural equation modeling is performed on aggregate data and provides results that are informative about populations, its results in this case cannot be used to obtain useful or reliable individual scores. A clearer understanding of the instrument’s psychometric structure will be of use both to researchers and to faculty tasked with remediating students who may have difficulty in this area.

Methods
In total, 4672 students from 30 osteopathic medical schools who took the COMLEX-USA Level 2-PE for the first time between July 7, 2014, and April 30, 2015, were included in the initial sample pool. Students provided informed consent to have their archival data used for research purposes when they registered for the COMLEX-USA series of examinations. Institutional review board approval for the present study was obtained from the Center for the Advancement of Healthcare Education and Delivery in Colorado Springs, Colorado.

The final sample consisted of 1 encounter per student, randomly sampled from the 3 to 4 OMT-scored encounters in the examination. One encounter per student was included to prevent statistical dependencies from clouding the results of the analyses.

OMT Scoring on the COMLEX-USA Level 2-PE
The COMLEX-USA Level 2-PE integrates osteopathic principles and practice throughout. The examination consists of 12 objective structured clinical examination stations through which candidates rotate during the testing day. In 25% to 40% of those encounters, candidates are scored on their use of the skills required for the safe and effective performance of OMT. Trained osteopathic physician raters (ie, OMT raters) watch the encounters remotely on a secure Web-based portal and assign scores to candidate performances in accordance with the OMT Global Rating Tool.

The rating tool has 6 components: (1) OMT modality selected, (2) OMT technique, (3) osteopathic examina-
tion and evaluation, (4) patient-physician position for treatment, (5) posttreatment assessment, and (6) treatment repetition and duration. Each of these components is scored on a Likert-type scale of 0 to 8 and is based on a scoring rubric containing behavioral anchors. The resulting scores are combined with scores from Data Gathering (History Taking and Physical Examination, which is recorded by the standardized patient using a case-specific checklist) and subjective, objective, assessment, and plan notes (scored by physician examiners) to compute the candidate’s final score for the Biomedical/Biomechanical component of the examination. Candidates must pass this domain to pass the examination, but the skills within the domain are compensatory in computing a final pass or fail decision for the domain. Only data from the 6 OMT components were included in these analyses.

**Statistical Analysis**

Correlations among student scores on the 6 components of the OMT Global Rating Tool were computed as a preliminary analysis. The results from the correlations among variables at the manifest level were used to inform the specification of the structural models of the underlying latent constructs.

The dimensionality of the OMT scale was assessed using structural equation modeling in SAS/STAT software (SAS for Windows, version 9.4; SAS Institute). Osteopathic manipulative treatment data were analyzed separately from the rest of the examination. The model was identified by constraining the variance of the latent variables to 1.0. No correlated error was allowed among indicator variables (ie, OMT assessment components).

The fit of the model to the data was evaluated using the Standardized Root Mean Residual (SRMR), a measure of the degree to which variance in the data remains unaccounted for by the model, and the Comparative Fit Index (CFI), which indicates the degree to which the specified model is an improvement over a model in which all variables are uncorrelated. The improvement in fit between competing models was tested using the difference in the 2 obtained $\chi^2$ statistics, evaluated as $\chi^2_i$:

A criterion of SRMR less than or equal to 0.06 and CFI greater than or equal to 0.96 was used to evaluate whether the model fit the data well.\textsuperscript{5}

**Results**

Data consisted of component-level scores from 4672 third- and fourth-year osteopathic medical students who took the COMLEX-USA Level 2-PE between July 7, 2014, and April 30, 2015. Of the students for whom demographic information was available, 2948 of 4398 were white (67.0%) and 2428 of 4397 were men (55.2%). This student demographic is consistent with the 2014 matriculating class of osteopathic medical students, which comprises 64.9% white students and 56.8% men.\textsuperscript{6}

The correlations among student raw scores are presented in Table 1. The values of these correlations indicate that the components related to OMT technique were strongly interrelated ($r=-0.57$ to $0.73$) and the 2 components related to pre– and post–OMT technique assessment are strongly interrelated ($r=0.44$), but the relationships between the 2 component sets were not as strong ($r=0.18$ to $0.34$). These results were used to guide subsequent model-building after the fit of the data to a 1-factor model proved unsatisfactory. The 1-factor model fit the data poorly ($\chi^2=1072.07$, $P<.001$, SRMR=0.078, CFI=0.91), indicating that the OMT Global Rating Tool is unlikely to be measuring 1 monolithic latent skill. On examination of the pattern of correlations among the indicator variables (Table 2), it seemed both statistically and conceptually reasonable to fit a 2-factor model, which significantly improved the overall model fit ($\Delta\chi^2=689.65$, $P<.001$, SRMR=0.03, CFI=0.97).

The 2-factor model (Figure) comprises a factor representing pre– and post–OMT technique assessment and a factor representing the student’s actual use of OMT techniques. Two components of the OMT scoring tool (osteopathic examination and evaluation and posttreatment assessment) were constrained to load on the first factor and 4 components of the OMT scoring tool (OMT modality selected, OMT technique, patient-physician position for treatment, patient-physician position for treatment, and posttreatment assessment) were constrained to load on the second factor, which was named OMT technique, patient-physician position for treatment, and posttreatment assessment.
position for treatment, and treatment repetition and duration) were constrained to load on the second factor. The correlation between latent variables (ie, the 2 factors) was 0.47; Wald tests indicated that this parameter could not be constrained to 0 without significantly decreasing model fit. A correlation of 0.47 between the latent variables, though not trivial, indicates a substantial amount of variance unique to each construct. Although the ideal conditions for model estimation call for 3 manifest variables per latent variable, in this case the path between the latent variable and the constraint of the variance of the latent variables to 1.0 was considered sufficient to ensure the identification of the model. However, the identification of the model depends on the ability of the 2-indicator pre– and post–OMT technique assessment factor to borrow strength from the OMT technique factor; the pre– and post–OMT technique assessment factor cannot be estimated separately.

The standardized solution for the 2-factor model is presented in Table 2 and gives the correlation between each OMT scoring rubric component and its respective latent variable.

<table>
<thead>
<tr>
<th>OMT Assessment Components</th>
<th>OMT Modality Selected</th>
<th>OMT Technique</th>
<th>Osteopathic Examination and Evaluation</th>
<th>Patient-Physician Position for Treatment</th>
<th>Posttreatment Assessment</th>
<th>Treatment Repetition and Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMT Modality Selected</td>
<td>r</td>
<td>0.67</td>
<td>0.20</td>
<td>0.68</td>
<td>0.26</td>
<td>0.62</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>4654</td>
<td>4662</td>
<td>4670</td>
<td>4663</td>
<td>4670</td>
</tr>
<tr>
<td>OMT Technique</td>
<td>r</td>
<td>...</td>
<td>0.25</td>
<td>0.70</td>
<td>0.30</td>
<td>0.73</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>...</td>
<td>4647</td>
<td>4655</td>
<td>4648</td>
<td>4655</td>
</tr>
<tr>
<td>Osteopathic Examination and Evaluation</td>
<td>r</td>
<td>...</td>
<td>...</td>
<td>0.18</td>
<td>0.44</td>
<td>0.27</td>
</tr>
<tr>
<td>N</td>
<td>...</td>
<td>...</td>
<td>4662</td>
<td>4659</td>
<td>4662</td>
<td></td>
</tr>
<tr>
<td>Patient-Physician Position for Treatment</td>
<td>r</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.22</td>
<td>0.57</td>
</tr>
<tr>
<td>N</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>4663</td>
<td>4670</td>
<td></td>
</tr>
<tr>
<td>Posttreatment Assessment</td>
<td>r</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.34</td>
</tr>
<tr>
<td>N</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>4663</td>
<td></td>
</tr>
<tr>
<td>Treatment Repetition and Duration</td>
<td>r</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Raw Score, mean (SD)</td>
<td>5.76 (1.85)</td>
<td>4.79 (1.78)</td>
<td>3.93 (1.89)</td>
<td>5.55 (1.90)</td>
<td>2.00 (1.97)</td>
<td>4.73 (1.88)</td>
</tr>
</tbody>
</table>

*P<.01 for all correlations. Individual N counts are based on number of cases in the data set for which each rating pair were available. Osteopathic examination and evaluation and posttreatment assessment comprised the “pre– and post–OMT (osteopathic manipulative treatment) technique assessment” latent variable. OMT modality selected, OMT technique, patient-physician position for treatment, and treatment repetition and duration comprised the “OMT technique” latent variable.
A student who is skilled at preassessment evaluation is generally, though not invariably, skilled at postassessment intervention as well; similarly, a student who is good at selecting the correct treatment modality can usually, though not invariably, do a reasonably good job at performing the treatment as well. The evidence of this model indicates, however, that a good preassessment evaluation does not necessarily lead to appropriate or skilled technique. In addition, a student who is skilled in performing OMT techniques is not necessarily good at completing thorough evaluations beforehand or assessments afterward. As measured by this instrument, assessment and technique appear to be distinct, though correlated, skill sets.

Many curricular models for teaching OMT at osteopathic medical schools exist, and the current study’s findings suggest a disconnect (as demonstrated in the difference in scores) in students’ skills in the diagnosis of somatic dysfunction vs the performance of an OMT technique. Incorporating OMT into the third- and fourth-year curriculum, when students can integrate diagnosis with treatment, might be helpful. Students who have clinical exposure to OMT before entering clinical training are more likely to plan to use OMT in future practice. Perhaps practicing OMT in clinical settings would help to integrate diagnosis and treatment. Additional studies would be helpful.

A limitation of the present study is that some students were not rated on some OMT components, which led to missing data that cannot be considered random. However, encounters with fewer than 6 scores accounted for 0.66% of the sample. A more immediate limitation is that the data analyses were performed on 1 cohort of students. Changes in curriculum, cohort ability level, or the level of student interest in OMT as a foundational skill can all influence the relationship between the latent traits measured by the OMT Global Rating Tool, or the relationship between the indicators and their corresponding latent variables.

### Table 2. Correlation Between Indicator Variables (OMT Assessment Components) and Their Respective Latent Variables Using a 2-Factor Model

<table>
<thead>
<tr>
<th>OMT Assessment Components</th>
<th>Pre- and Post-OMT Technique Assessment</th>
<th>OMT Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMT modality selected</td>
<td>0</td>
<td>0.79&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>OMT technique</td>
<td>0</td>
<td>0.89&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Osteopathic examination and evaluation</td>
<td>0.60&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Patient-physician position for treatment</td>
<td>0</td>
<td>0.79&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Posttreatment assessment</td>
<td>0.72&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Treatment repetition and duration</td>
<td>0</td>
<td>0.80&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Latent variables were grouped as follows: pre- and post–osteopathic manipulative treatment (OMT) technique assessment included osteopathic structural examination and evaluation and posttreatment assessment, and OMT technique included OMT modality selected, OMT technique, position for treatment, and treatment repetition and duration.

<sup>b</sup> P < .01.

### Discussion

The correlations among the 6 components of the OMT Global Rating Tool, viewed at the raw score level, suggested 2 clusters of measured skills: 1 related to pre- and post–OMT technique assessment and 1 related to skill at the actual OMT technique. When structural equation modeling was used to test the hypothesis that the tool comprises a single global OMT skill, the hypothesis was rejected because the model was inconsistent with the observed relationships among student component scores. A model that separated the 2 groups of skills fit the data better and was not rejected on statistical grounds.

These results suggest that skill at OMT as measured by the NBOME’s OMT Global Rating Tool is a set of 2 distinct but moderately correlated skills: performing an osteopathic assessment of a patient and performance of a specific OMT technique. A student who is skilled at preassessment evaluation is generally, though not invariably, skilled at postassessment intervention as well; similarly, a student who is good at selecting the correct treatment modality can usually, though not invariably, do a reasonably good job at performing the treatment as well. The evidence of this model indicates, however, that a good preassessment evaluation does not necessarily lead to appropriate or skilled technique. In addition, a student who is skilled in performing OMT techniques is not necessarily good at completing thorough evaluations beforehand or assessments afterward. As measured by this instrument, assessment and technique appear to be distinct, though correlated, skill sets.

Many curricular models for teaching OMT at osteopathic medical schools exist, and the current study’s findings suggest a disconnect (as demonstrated in the difference in scores) in students’ skills in the diagnosis of somatic dysfunction vs the performance of an OMT technique. Incorporating OMT into the third- and fourth-year curriculum, when students can integrate diagnosis with treatment, might be helpful. Students who have clinical exposure to OMT before entering clinical training are more likely to plan to use OMT in future practice. Perhaps practicing OMT in clinical settings would help to integrate diagnosis and treatment. Additional studies would be helpful.

A limitation of the present study is that some students were not rated on some OMT components, which led to missing data that cannot be considered random. However, encounters with fewer than 6 scores accounted for 0.66% of the sample. A more immediate limitation is that the data analyses were performed on 1 cohort of students. Changes in curriculum, cohort ability level, or the level of student interest in OMT as a foundational skill can all influence the relationship between the latent traits measured by the OMT Global Rating Tool, or the relationship between the indicators and their corresponding latent variables.
Conclusion
The OMT assessment rubric used by the NBOME for the COMLEX-USA Level 2-PE appears to measure 2 distinct, moderately correlated constructs—(1) the skill involved in identifying and performing the most appropriate technique for the patient presentation and (2) the skill involved in pre– and post–OMT technique assessment of the patient’s somatic dysfunction. A correlation of 0.47 between the latent variables, though not trivial, indicates a substantial amount of variance unique to each construct; a substantial amount of skill unique to pre– and post–OMT technique assessment is unrelated to skill at OMT technique, and vice versa. Although latent-specific subscores cannot be reliably provided by this instrument, it might nonetheless be helpful to remediation efforts to consider pre– and post–OMT technique assessments and OMT techniques to be distinct areas to target in interventions.

Author Contributions
Drs Smith and Sandella and Ms Xu provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; Drs Smith, Sandella, and Dowling drafted the article or revised it critically for important intellectual content; Drs Smith and Sandella gave final approval of the version of the article to be published; and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

© 2016 American Osteopathic Association

Figure
Illustration of the 2-factor model that comprises a factor representing pre– and post–OMT technique assessment and a factor representing the student’s actual use of OMT techniques.