Relationship of Admissions Variables and College of Osteopathic Medicine Variables to Performance on COMLEX-USA Level 3

Helen H. Baker, PhD, MBA; Victoria L. Shuman, DO; Lance C. Ridpath, MS; Lorenzo L. Pence, DO; Robert M. Fisk Jr, PhD, MA; and Craig S. Boisvert, DO

Context: New accreditation standards require that all US colleges of osteopathic medicine (COMs) publically report the first-time pass rates of graduates on the Comprehensive Osteopathic Medical Licensing Examination-USA (COMLEX-USA) Level 3. Little is known about the extent to which admissions variables or COM performance measures relate to Level 3 performance.

Objective: To examine the relationship of admissions variables and COM performance to scores on Level 3 and to assess whether a relationship existed between Level 3 scores and sex, curriculum track, year of graduation, and residency specialty in the first postgraduate year.

Methods: Data were analyzed from 4 graduating classes (2008-2011) of the West Virginia School of Osteopathic Medicine in Lewisburg. Relationships were examined between first-attempt scores on COMLEX-USA Level 3 and Medical College Admission Test (MCAT) scores; undergraduate grade point averages (GPAs); GPAs in COM year 1, year 2, and clinical rotation years (years 3 and 4); and first-attempt scores on COMLEX-USA Level 1, Level 2-Cognitive Evaluation, and Level 2-Performance Evaluation.

Results: Of the 556 graduates during this 4-year period, COMLEX-USA Level 3 scores were available for 552 graduates (99.3%). No statistically significant differences were found in Level 3 scores based on sex, curriculum track, graduating class, or residency specialty. The strongest relationship between Level 3 scores and any admissions variable was with total MCAT score, which accounted for 4.2% of the variation in Level 3 scores. The strongest relationship between Level 3 scores and COM year performance measures was with year 2 GPA, which accounted for 35.4% of the variation in Level 3 scores. Level 1 scores accounted for 38.5% of the variation in Level 3 scores, and Level 2-Cognitive Evaluation scores accounted for the greatest percentage of variation (45.7%). The correlation of Level 3 scores with passing the Level 2-Performance Evaluation on the first attempt was not statistically significant.

Conclusion: A weak relationship was found between admissions variables and performance on COMLEX-USA Level 3, suggesting that graduates with lower MCAT scores and undergraduate GPAs may have overcome their early disadvantage. Strong relationships were found between Level 3 scores and year 2 GPAs, as well as scores on COMLEX-USA Level 1 and Level 2-Cognitive Evaluation.
The relationship between COM performance and performance on components of Levels 1 or 2. Studies have also examined the relationships among performance on the various levels of COMLEX-USA, including Level 3, and an increasing number of studies have shown a relationship between Level 3 performance and other measures of performance in and after residency. Multiple studies on the United States Medical Licensing Examination (USMLE) Step 3, the analogous examination used by MD degree–granting medical schools, have found that performance on USMLE Step 3 depended in part on whether graduates sought a broad-based residency or one with a more narrow focus. However, no studies could be identified that correlated COMLEX-USA Level 3 performance to admissions variables or performance in the COM curriculum other than to scores on Levels 1 and 2. If performance on Level 3 is to be used in accreditation decisions about COMs, and if admissions variables are at some point to be regarded as relating to Level 3 score, then these relationships need to be elucidated.

The current study was designed to examine the relationship of admissions variables and COM performance to performance on Level 3. Variables included undergraduate grade point averages (GPAs), Medical College Admission Test (MCAT) scores, COM GPAs (year 1, year 2, and clinical rotation [years 3 and 4]), and COMLEX-USA Level 1, Level 2-CE, and Level 2-PE scores. In addition, we sought to assess whether a relationship existed between Level 3 scores and sex, curriculum track, year of graduation, and residency specialty in the first postgraduate year.

Methods

Participants and Setting

The mission of the West Virginia School of Osteopathic Medicine (WVSOM) is focused on preparing graduates to provide primary care in rural areas of West Virginia. Data from all WVSOM graduates (classes 2008-2011) were initially included, even if a
The COM performance variables included GPA at the end of year 1, year 2, and the clinical rotation years. At WVSOM, grades are on a scale from 0 to 100, with a score of 70 required for passing. Grade point averages were calculated according to the procedures used for student transcripts. The other COM variable was COMLEX-USA Level 1 and 2 scores. The primary path to licensure for osteopathic physicians, the COMLEX-USA, has 3 levels. Level 1 is an 8-hour computer-based examination completed at the end of COM year 2. Level 2 is taken during the clinical rotations and has 2 components: the Level 2-CE, an 8-hour computer-based examination, and the Level 2-PE, a 1-day performance examination in which each candidate takes part in 12 standardized patient encounters. If a WVSOM student had multiple scores on a level or component of the COMLEX-USA, the score obtained on the first attempt was used. Level 3 is currently an 8-hour computer-based examination of knowledge obtained during residency training.

Based on the residency program that graduates entered for their first postgraduate year, graduates were assigned to 1 of 3 categories: broad-based (family practice, internal medicine, pediatrics, emergency medicine, or any combination of these specialties); traditional internship or transitional year; and not broad-based (all other residencies).

We anticipated that too few graduates would fail Level 3 to allow meaningful analysis based on failures; therefore, we looked at performance on Level 3 as a continuous variable (3-digit scores). For analysis of differences based on sex and curriculum track, an independent 2-sample t test was used. An analysis of variance was used for differences between year of graduation and residency specialty. For all continuous variables—admissions variables, COM GPAs, and Level 1 and Level 2-CE scores—we calculated Pearson product moment correlations with Level 3 scores and statistical significance. We also squared the correlation to obtain the percentage of variation in Level 3 scores that student had initially failed a program requirement, such as a course or Level 1 or 2 of COMLEX-USA.

The majority of WVSOM’s students in the classes studied had been enrolled in a systems-based curriculum. A smaller number of students had been enrolled in a modified problem-based learning curriculum, which was implemented with the graduating class of 2007 and discontinued with the class of 2013. For the purposes of this study, a student’s curriculum track was designated at the time of matriculation. Students in both curriculum tracks completed the same curriculum in the clinical rotations.

The clinical curriculum at WVSOM consists primarily of rotations in ambulatory and hospital settings. Grades in WVSOM’s clinical curriculum are determined primarily by preceptor ratings (ranging from 70% to 100% of the grade), with some rotations requiring postrotation written tests or other requirements, such as logs or graded case studies. The school conducts Objective Structured Clinical Evaluations, but for the graduates included in this analysis, performance on these evaluations was not a factor in any GPA. Furthermore, for the years studied, no score on any test by the National Board of Osteopathic Medical Examiners was part of any GPA. The class size increased with the class that graduated in 2010, so we also looked for differences in Level 3 scores based on year of graduation.

The present study was exempt from human subjects review by the institutional review board at WVSOM. Analyses were performed using the SAS 9.2 statistical software program (SAS Institute, Inc).

Variables and Analysis
Admissions variables included undergraduate science, non-science, and total GPAs; MCAT Verbal Reasoning, Physical Sciences, and Biological Sciences subtest scores; and total MCAT score. For applicants who had taken the MCAT more than once, the highest total MCAT score and the highest subtest scores were used.

The COM performance variables included GPA at the end of year 1, year 2, and the clinical rotation years. At WVSOM, grades are on a scale from 0 to 100, with a score of 70 required for passing. Grade point averages were calculated according to the procedures used for student transcripts. The other COM variable was COMLEX-USA Level 1 and 2 scores. The primary path to licensure for osteopathic physicians, the COMLEX-USA, has 3 levels. Level 1 is an 8-hour computer-based examination completed at the end of COM year 2. Level 2 is taken during the clinical rotations and has 2 components: the Level 2-CE, an 8-hour computer-based examination, and the Level 2-PE, a 1-day performance examination in which each candidate takes part in 12 standardized patient encounters. If a WVSOM student had multiple scores on a level or component of the COMLEX-USA, the score obtained on the first attempt was used. Level 3 is currently an 8-hour computer-based examination of knowledge obtained during residency training.

Based on the residency program that graduates entered for their first postgraduate year, graduates were assigned to 1 of 3 categories: broad-based (family practice, internal medicine, pediatrics, emergency medicine, or any combination of these specialties); traditional internship or transitional year; and not broad-based (all other residencies).

We anticipated that too few graduates would fail Level 3 to allow meaningful analysis based on failures; therefore, we looked at performance on Level 3 as a continuous variable (3-digit scores). For analysis of differences based on sex and curriculum track, an independent 2-sample t test was used. An analysis of variance was used for differences between year of graduation and residency specialty. For all continuous variables—admissions variables, COM GPAs, and Level 1 and Level 2-CE scores—we calculated Pearson product moment correlations with Level 3 scores and statistical significance. We also squared the correlation to obtain the percentage of variation in Level 3 scores that
could be accounted for by knowing the other variable. As Level 2-PE is reported to COMs only as pass/fail and, therefore, is a dichotomous variable, we used an independent 2-sample t test.

Because we ran multiple significance tests, a Bonferroni correction was applied. Using the standard type I error rate of \( \alpha=0.05 \), the Bonferroni adjustment created an alternate P value cutoff of \( \alpha_{\text{adjusted}} = 0.0029 \), where \( \alpha_{\text{adjusted}} = 0.05/17 \), because 17 statistical tests were performed.

**Results**

During the 4-year study period (2008-2011), WVSOM had 556 graduates. As of April 1, 2014, Level 3 scores were available for 552 graduates (99.3%). A total of 4 students were excluded from the data analysis (3 from the class of 2010 and 1 from the class of 2011) because their Level 3 scores were not available at the time of analysis. Of the 552 graduates with a Level 3 score, 521 (94.4%) passed the examination on their first attempt, and all of those who failed on the first attempt (31 [100%]) eventually passed.

*Table 1* presents the characteristics of the sample. No statistically significant differences were found in Level 3 performance based on sex, curriculum track, graduating class, or residency specialty.

Correlations of continuous variables with COMLEX-USA Level 3 are shown in *Table 2*, and the *Figure* shows the percentage of variation in Level 3 scores accounted for by these variables. Most correlations between these variables and Level 3 scores were statistically significant, although they varied in practical value. The strongest correlation between Level 3 and any admissions variable was found with the total MCAT score \( (r^2=0.205) \), accounting for 4.2% (the correlation squared) of the variation in Level 3. Of performance measures in COMs, the greatest percentage of variation in Level 3 was accounted for by year 2 GPA, which accounted for 35.4% of the variation in Level 3. Level 1 scores accounted for 38.5% of the variation in Level 3 scores, and Level 2-CE scores provided the greatest percentage of variation (45.7%).

The mean (SD) Level 3 score for the 14 students who had initially failed the Level 2-PE was 488.5 (124.6), and the mean score for the 538 students initially passing the Level 2-PE was 517.7 (112.2). An independent 2-sample t test showed that this difference was not statistically significant \( (P=0.338) \).

**Discussion**

To our knowledge, no studies have documented the degree of relationship between admissions variables or COM GPA and COMLEX-USA Level 3 scores. The current study found weak relationships between admissions variables and Level 3 performance, with the strongest relationship (total MCAT score) accounting for less than 5% of the variation in Level 3 performance. The admissions data identified in other studies as predicting success in year 1 and year 2 and success on Level 1 were of little help in predicting performance on Level 3. An optimistic view of these findings is that students who have relatively low admissions statistics but successfully graduate have overcome their early disadvantage, and on their first attempt attained Level 3 scores that approximate those of students who had better admissions statistics. Furthermore, the relationship of Level 3 scores with admissions variables was weak, indicating that the AOA COCA requirement of a college to “tie its admissions process and criteria to the outcome performance of its graduates” should perhaps be revisited in the future. The findings of the present study suggest that AOA COCA reviewers should continue to focus on whether a college admits students who “have the ability to complete the curriculum” rather than extending this standard to apply to performance during residency.

Regarding the relationship between performance in the COM curriculum and performance on Level 3, the year 2 GPA had the strongest relationship, accounting for 35.4% of the variation in Level 3 scores. Therefore, year 2 GPA seems to be an appropriate
researchers in Canada found a relationship between scores on Canada’s licensing examination taken in residency and patient outcomes.29-32 On the basis of these findings, more research is needed.

Regarding clinical rotation GPA, ratings by preceptors comprise 70% to 100% of WVSOM’s GPAs on clinical rotations, so clinical rotation GPA is less a measure of knowledge (as it would be measured on a multiple-choice test) than a measure of clinical performance. Therefore, like the Level 2-PE score, clinical ro-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level 3 Score</th>
<th>No. (%)</th>
<th>Mean (SD)</th>
<th>P Value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>271 (49)</td>
<td>513 (105)</td>
<td>.42$^b$</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>281 (51)</td>
<td>521 (119)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum Track</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems-based learning</td>
<td>449 (81)</td>
<td>521 (113)</td>
<td>.05$^b$</td>
<td></td>
</tr>
<tr>
<td>Modified problem-based learning</td>
<td>103 (19)</td>
<td>498 (109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduating Class</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>96 (17)</td>
<td>504 (126)</td>
<td>.15$^c$</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>100 (18)</td>
<td>500 (106)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>165 (30)</td>
<td>526 (103)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>191 (35)</td>
<td>525 (116)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency Specialty in First Postgraduate Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad-based categorical$^d$</td>
<td>351 (64)</td>
<td>521 (111)</td>
<td>.52$^c$</td>
<td></td>
</tr>
<tr>
<td>Traditional internship or transitional year</td>
<td>89 (16)</td>
<td>505 (109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not broad-based, categorical$^e$</td>
<td>112 (20)</td>
<td>515 (120)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Using the standard type I error rate of $\alpha=0.05$, the Bonferroni adjustment created an alternate $P$ value cutoff of an adjusted $\alpha$ of .0029, where the adjusted $\alpha$ was .05/17, because 17 statistical tests were performed in the analysis reported in this study.

$^b$ Independent 2-sample t test.

$^c$ Analysis of variance.

$^d$ Family practice, internal medicine, pediatrics, emergency medicine, or any combination of these 3 alternatives.

$^e$ All residencies other than those categorized as broad-based categorical or as traditional internship or transitional year.

Abbreviation: COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA.
2-CE scores. Poor performance on Level 1 or Level 2-CE seems to be a “red flag” for poor performance on Level 3. The relationship between Level 2-PE and Level 3 scores was small and not statistically significant. This finding, too, matched findings in the literature, suggesting that the Level 2-PE may tap different skills from those assessed by computer-based examinations.

Of the graduates with a score reported, 521 (94.4%) passed the Level 3 examination on their first attempt, suggesting that GPA should not be expected to correlate strongly with the Level 3 knowledge examination score. Future research might specifically examine the clinical core competencies being taught and assessed in both predoctoral and residency training, including the ability to perform osteopathic manipulative treatment and comparing these subscores with Level 3 performance.

The present study found correlations greater than 0.60 between Level 3 scores and both the Level 1 and the Level 2-CE scores. Poor performance on Level 1 or Level 2-CE seems to be a “red flag” for poor performance on Level 3.

Table 2. Correlations of Continuous Variables With COMLEX-USA Level 3 Scores for the Graduating Classes of 2008 to 2011 at the West Virginia School of Osteopathic Medicine (N=552)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation With Level 3</th>
<th>P Value</th>
<th>% of Variation Accounted fora</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admissions Variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonscience</td>
<td>0.097</td>
<td>.023*</td>
<td>0.9</td>
</tr>
<tr>
<td>Science</td>
<td>0.160</td>
<td>&lt;.001</td>
<td>2.6</td>
</tr>
<tr>
<td>Total</td>
<td>0.152</td>
<td>&lt;.001</td>
<td>2.3</td>
</tr>
<tr>
<td>MCAT score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>0.138</td>
<td>.001</td>
<td>1.9</td>
</tr>
<tr>
<td>Physical sciences</td>
<td>0.111</td>
<td>.009*</td>
<td>1.2</td>
</tr>
<tr>
<td>Biological sciences</td>
<td>0.182</td>
<td>&lt;.001</td>
<td>3.3</td>
</tr>
<tr>
<td>Total</td>
<td>0.205</td>
<td>&lt;.001</td>
<td>4.2</td>
</tr>
<tr>
<td>Performance in the COM Curriculum, GPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>0.462</td>
<td>&lt;.001</td>
<td>21.4</td>
</tr>
<tr>
<td>Year 2</td>
<td>0.595</td>
<td>&lt;.001</td>
<td>35.4</td>
</tr>
<tr>
<td>Clinical years</td>
<td>0.312</td>
<td>&lt;.001</td>
<td>9.8</td>
</tr>
<tr>
<td>COMLEX-USA Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>0.620</td>
<td>&lt;.001</td>
<td>38.5</td>
</tr>
<tr>
<td>Level 2-CE</td>
<td>0.676</td>
<td>&lt;.001</td>
<td>45.7</td>
</tr>
</tbody>
</table>

a Correlation squared.
b Not statistically significant. Using the standard type I error rate of α=.05, the Bonferroni adjustment created an alternate P value cutoff of an adjusted α of .0029, where the adjusted α was .05/17, because 17 statistical tests were performed in the analysis reported in this study.

Abbreviations: COM, college of osteopathic medicine; COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA; GPA, grade point average; MCAT, Medical College Admission Test.
reflecting the national percentage of graduates who passed Level 3 on the first attempt between 2007 and 2012 (93%).33,34 (The national average increased to 95.1% in the 2012-2013 academic year.)

A key question is the extent to which performance on Level 3 should be attributed to a resident’s COM or to the graduate medical education program. The AOA COCA standard requiring colleges to report first-time pass rates on Level 3 implies a belief that COMs should be held accountable for this performance; however, postgraduate education is a time of tremendous learning. Colleges of osteopathic medicine should be held accountable for their students’ scores on Level 1, Level 2-CE, and Level 2-PE, but it is our opinion that residency programs should share the responsibility for scores on Level 3. In the current study, the strongest correlation with performance on Level 3 was Level 2-CE scores, which accounted for about half the variation in Level 3 scores, leaving half the variation in Level 3 scores not accounted for. A multisite study might analyze Level 3 scores while controlling for both Level 2-CE performance and residency program type to tease out the amount of variation in Level 3 scores that would be attributed to COMs.

The methodology in the present study was slightly different from that used in other studies, in that we...
included every graduate from these classes for whom a first attempt score on Level 3 was available. Studies such as that by Dixon included only those students who had graduated in 4 years. Also, we used the highest total MCAT score and the subscores associated with that attempt rather than the highest subscore obtained, as some others have done (eg, Evans et al). Researchers using other methods might obtain different results.

The modified problem-based learning program that was briefly in effect at WVSOM was substantially different from the authentic problem-based learning program as described by Barrows and Tamblyn, in that WVSOM’s students in this curriculum track spent large portions of their time in traditional learning activities. Therefore, no conclusions should be drawn from the current study to the larger literature on problem-based learning.

The graduates in the current study were from a single, unique COM focused on educating students for service in West Virginia in primary care and rural medicine. Studies have identified WVSOM as having the highest percentage of graduates in rural practice of any MD- or DO-granting medical school in the nation and as having the largest number of graduates in rural practice in Appalachia. Furthermore, more than 65% of recent WVSOM graduates initially entered primary care residencies. The admissions process at WVSOM and the curriculum are focused on these outcomes and may not be typical of other COMs.

We did not attempt to quantify the admissions committee’s assessment of the attitudes, values, and career plans of applicants, information that the WVSOM faculty use as part of the “holistic review” of applicants, data that may be related to Level 3 scores. Furthermore, the only residency measure included in the current study was the Level 3 score. Future studies of residency performance and Level 3 should include direct measures of clinical competence or patient outcomes data.

Changes in the assessment of potential and current osteopathic medical students are anticipated at a national level. Revision of the MCAT is planned, which may result in an examination that better predicts performance in osteopathic medical school and in residency. Regarding COMLEX-USA Level 3, changes are under development and being pilot tested for adoption, with a target date of 2017 or 2018. We expect the current knowledge component to remain, and a second day of testing to be added to assess core competencies that are not well measured by multiple-choice tests (personal communication, B. Bates, DO, 2014). This new component of Level 3 might have a stronger relationship with performance-based examinations. Further research will be needed on both of these new examinations.

Conclusion
The current study at WVSOM found little relationship between admissions variables and performance on COMLEX-USA Level 3. A strong relationship was found between performance on Level 3 and the COM year 2 GPA and Level 1 and Level 2-CE scores. Performance on Level 2-PE was not related to performance on Level 3.

Acknowledgments
We thank the following individuals for their thoughtful suggestions on earlier drafts of this manuscript: Bruce Bates, DO; Hugh Clements-Jewery, PhD; and Brian Griffith, PhD. We are also grateful for the assistance in manuscript preparation provided by Bethany M. Hampton, BS.

Author Contributions
All authors provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; all authors drafted the article or revised it critically for important intellectual content; and Drs Baker, Boisvert, Fisk, and Pence gave final approval of the version of the article to be published.

(continued)
References


