

Relationships Between the Comprehensive Osteopathic Medical Achievement Test (COMAT) Subject Examinations and the COMLEX-USA Level 2-Cognitive Evaluation

Feiming Li, PhD; Kevin E. Kalinowski, PhD; Hao Song, PhD; and Bruce P. Bates, DO

From the Department of Cognitive Testing at the National Board of Osteopathic Medical Examiners (NBOME) in Chicago, Illinois.

Financial Disclosures: All authors are employees of the NBOME.

Support: This study was funded by the NBOME.

Address correspondence to Kevin E. Kalinowski, PhD, Psychometrician, National Board of Osteopathic Medical Examiners, 8765 W Higgins Rd, Suite 200, Chicago, IL 60631-4174.

E-mail: kkalinowski@nbome.org

Submitted October 1, 2013; final revision received March 10, 2014; accepted April 21, 2014.

Context: The relationship between the Comprehensive Osteopathic Medical Achievement Test (COMAT) series of subject examinations and the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Cognitive Evaluation (COMLEX-USA Level 2-CE) has not been thoroughly examined.

Objective: To investigate the factors associated with performance on COMAT subject examinations and how COMAT scores correlate with COMLEX-USA Level 2-CE scores.

Methods: We examined scores of participants from 2 COMAT examination cycles in 2011 and 2012. According to surveys, most schools used COMAT scores in clerkship and clinical rotation evaluation, which were classified as being used for “high-stakes” purposes. We matched first-attempt COMAT scores with first-attempt COMLEX-USA Level 2-CE scores, and we conducted correlation analyses between the scores from the 7 COMAT subject examinations, as well as between COMAT and COMLEX-USA Level 2-CE scores. Multiple linear regression analyses were performed to investigate how much variance in COMLEX-USA Level 2-CE scores was explained by COMAT scores.

Results: In 2011 and 2012, respectively, 3751 and 3786 COMAT candidates had COMLEX-USA Level 2-CE scores (53.0% and 93.9%, respectively, had ≥ 1 high-stakes COMAT score). Intercorrelations between COMAT scores were low to moderate ($r=0.27-0.53$), as hypothesized. Correlations between COMAT and Level 2-CE scores were moderate to high, with the highest correlations for internal medicine COMAT scores ($r=0.63-0.65$). All regressions showed internal medicine scores as the strongest predictor of Level 2-CE performance. Groups with high-stakes scores had larger adjusted coefficients of determination than those with low-stakes scores (eg, $R^2=0.63$ vs 0.52 , respectively, in 2011). For 2012 candidates with high-stakes scores, all predictors were statistically significant.

Conclusion: The COMAT subject examination scores were moderately inter-correlated, as hypothesized, with higher correlations between COMAT and COMLEX-USA Level 2-CE scores. The COMAT performance was predictive of COMLEX-USA Level 2-CE performance.

J Am Osteopath Assoc. 2014;114(9):714-721

doi:10.7556/jaoa.2014.140

During years 3 and 4 of osteopathic medical school, in addition to coursework and licensure examinations, students typically move through various clinical clerkships and rotations to gain patient-centered experience as practitioners of osteopathic medicine. In the past, most osteopathic medical schools used postrotation objective examinations and global rating tools completed by clinical preceptors to determine clerkship grades and criteria for academic promotion, but residency program directors expressed a need for standardized end-of-clerkship rotation examinations that were secure, integrated osteopathic principles, and provided direct evidence of student learning, along with some measure of educational equivalency at often-varied clinical education sites. In response, the National Board of Osteopathic Medical Examiners (NBOME) developed and began administration of the Comprehensive Osteopathic Medical Achievement Test (COMAT) series of subject examinations in 2011. These COMAT examinations were the first osteopathic achievement tests to feature not only examination blueprints, but also learning objectives and recommended teaching and learning resources. Currently, the NBOME offers COMAT examinations in the following 7 disciplines: family medicine, internal medicine, obstetrics-gynecology, osteopathic principles and practice (OPP), pediatrics, psychiatry, and surgery. A COMAT emergency medicine examination is expected to be released in 2015.

The Comprehensive Medical Licensing Examination (COMLEX-USA) is a series of assessments designed by the NBOME for the licensing of osteopathic physicians in the United States. The American Osteopathic Association Commission on Osteopathic College Accreditation requires that all osteopathic medical students pass the COMLEX-USA Level 1, Level 2-Cognitive Evaluation (CE), and Level 2-Performance Evaluation (PE) before graduation.¹ The Level 1 examination emphasizes the scientific concepts and principles necessary for understanding the mechanisms of health, clinical problems, and disease processes. Level 2-CE emphasizes the bio-

medical concepts and principles necessary for making appropriate clinical diagnoses based on patient history and physical examination findings. Level 2-PE is the associated standardized patient-based clinical skills component of the COMLEX-USA series. A passing score on these examinations indicates that the candidate's application of osteopathic medical knowledge and clinical skills have met an acceptable minimal competency standard for entry into graduate medical education (ie, graduation from osteopathic medical school). The ensuing final examination in the COMLEX-USA series (Level 3) is typically taken during the candidate's residency training.

The COMAT subject examinations primarily consist of single-best-answer questions in a clinical scenario format comparable to that of the COMLEX-USA series. At the time of the current study, COMAT examinations included 100 questions presented in a single 2-hour session for each content area. (Current COMAT examinations have 125 questions presented in 2½ hours.) The COMLEX-USA Level 2-CE includes 400 items delivered in 8 sections over 8 hours. Both series of examinations produce nationally standardized scores for reporting. The COMAT scores have a mean (SD) of 100 (10), and the COMLEX-USA Level 2-CE scores have a mean (SD) of 500 (100), with a passing score set at 400. As achievement tests, COMAT examinations do not have passing scores.

The NBOME initiated the current 3-level COMLEX-USA series beginning in 1995, and several studies have substantiated its validity and predictive capabilities.²⁻⁷ Because COMAT is a relatively new series of examinations, this is to our knowledge the first published study of it, making our findings especially important.

The purpose of the current study was to determine the interrelationship between performance on individual COMAT examinations and between COMAT and COMLEX-USA Level 2-CE performance. On the basis of seminal research in medical education,⁸ researchers have posited that achievement in a given domain, or subject area, is not entirely explainable by the transfer of

some general knowledge between domains.⁹ Rather, performance is content specific, and a physician's knowledge and skills are clustered within a specific domain. Therefore, performance in one subject area would not necessarily strongly predict performance in another distinct subject area. Research findings supports this theory¹⁰ but also suggest that transferable skills may be important in explaining related performance.¹¹ Consequently, a measure of a distinct subject area should have a stronger relationship to a comprehensive measure than to a measure of an associated subject.

Given the theoretical framework guiding our study, we proposed the following 4 hypotheses:

- COMAT subject examination scores will be intercorrelated, as well as positively correlated with COMLEX-USA Level 2-CE scores.
- Intercorrelations between COMAT subject examination scores will be lower than the correlations between COMAT subject examination scores and the COMLEX-USA Level 2-CE scores.
- Correlations between COMAT and COMLEX-USA Level 2-CE scores will be stable from one testing cycle to another.
- COMAT scores used in the evaluation of clerkship or clinical rotation performance will be stronger predictors of COMLEX-USA Level 2-CE scores than COMAT scores not used for these purposes.

Methods

For the current study we selected participants from 2 COMAT cycles in 2011 and 2012. The 2011 cycle, the initial cycle for COMAT, extended from July 1, 2011, to June 30, 2012, and the 2012 cycle from July 1, 2012, to June 30, 2013. According to surveys conducted by the NBOME in 2011 and 2012, most osteopathic medical schools used COMAT scores either as part of students' final grade or for a pass-fail decision in their clerkships and clinical rotations. However, a few schools used COMAT only for formative purposes and did not ascribe COMAT scores to actual performance; this applied to 39.6% of the COMAT administrations in the 2011 cycle. For the 2012 cycle, a larger proportion of schools used COMAT scores to assess performance, with only 8.3% of the examinations used formatively. For the current study, we classified the scores from schools who used COMAT for clerkship or clinical rotation grades or pass-fail determination as being used for "high-stakes" purposes, and we classified the rest as being used for "low-stakes" purposes. We used only existing data that were properly obtained; participants could not be identified, directly or statistically. Therefore, the research was exempt from institutional review board approval.

After first-attempt COMAT score records were matched with first-attempt COMLEX-USA Level 2-CE scores, we conducted correlation analyses between the scores from the 7 COMAT subject examinations and between COMAT and COMLEX-USA Level 2-CE scores for the following 4 groups: (1) 2011 low-stakes administrations; (2) 2011 high-stakes administrations; (3) 2012 low-stakes administrations; and (4) 2012 high-stakes administrations. We also performed a simultaneous multiple linear regression analysis of the groups to determine how much of the variance in COMLEX-USA Level 2-CE scores was explained by COMAT scores. All analyses were conducted using SAS/STAT software (SAS for Windows, version 9.3; SAS Institute). The *P* value cutoff (α) was set a priori to .05 for all tests of statistical significance.

Results

Table 1 lists the numbers of COMAT subject examination administrations, COMAT examinees, and COMAT examinees with matched COMLEX-USA Level 2-CE scores for 2011 and 2012. A total of 3897 examinees from 22 colleges of osteopathic medicine and 5625 from 24 colleges of osteopathic medicine took at least 1 COMAT subject examination in 2011 and 2012, respectively, and 2023 and 4640 examinees took at least 1 COMAT subject examination for high-stakes purposes in 2011 and 2012, respectively. Using NBOME identifiers, we matched first-attempt COMAT scores with a candidate's first-attempt COMLEX-USA Level 2-CE scores.

Typically, candidates take COMLEX-USA Level 2-CE in the year after their COMAT subject examinations. For our purposes, COMAT scores recorded after Level 2-CE scores or without matched Level 2-CE scores were dropped from the analysis. As a consequence, 3751 COMAT candidates in 2011 and 4069 in 2012 had Level 2-CE scores, with 1989 (53.0%) and 3803 (93.5%), respectively, having taken at least

1 COMAT subject examination for high-stakes purposes. Although both low- and high-stakes groups were included in subsequent analysis for the 2011 cycle, only the high-stakes group was included in the 2012 analysis, owing to the limited number of observations (n=266) in the 2012 low-stakes group.

Table 2 provides Pearson product-moment correlation results for the 7 high-stakes COMAT subject examinations in 2012. Intercorrelations between COMAT scores were moderate ($r=0.27-0.54$). Table 3 lists the correlations between COMAT and COMLEX-USA Level 2-CE scores for 2011 and 2012. Results indicated generally moderate to high correlations, plus a consistent pattern in relative strength of correlations in subjects from the 2011 to the 2012 testing cycle, showing that Level 2-CE scores had the highest correlation with scores for internal medicine ($r=0.62-0.65$), followed by surgery, family medicine, obstetrics-gynecology, and pediatrics ($r=0.52-0.61$) and then psychiatry ($r=0.50-0.52$), with the lowest correlation to scores for OPP ($r=0.38-0.45$).

Table 1.
Examination Administrations and Examinees by Group Classification in 2011 and 2012

Group ^a	COMAT, No.		No. Examinees With Matched COMLEX-USA Level 2-CE Scores
	Administrations	Examinees	
2011			
Low stakes	6302	1874	1762
High stakes	9599	2023	1989
Total	15,901	3897	3751
2012			
Low stakes	1837	985	266
High stakes	20,257	4640	3803
Total	22,094	5625	4069

^a A "high-stakes" examination was defined as a Comprehensive Osteopathic Medical Achievement Test (COMAT) in which the score was used for clerkship and clinical rotation evaluations. "Low-stakes" examination scores were not used for such purposes.

Abbreviation: COMLEX-USA Level 2-CE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Cognitive Evaluation.

Table 2.
Intercorrelations^a for the 7 High-Stakes^b COMAT Subject
Examinations Taken by Osteopathic Medical Students in 2012

Subject Examination	Obstetrics-Gynecology	OPP	Psychiatry	Pediatrics	Family Medicine	Internal Medicine	Surgery
Obstetrics-Gynecology							
<i>r</i>	...	0.33	0.47	0.54	0.50	0.53	0.52
<i>n</i>	...	1977	2382	2592	2286	2550	2785
OPP							
<i>r</i>	0.27	0.30	0.35	0.34	0.36
<i>n</i>	1925	1753	1484	1942	1954
Psychiatry							
<i>r</i>	0.45	0.43	0.47	0.43
<i>n</i>	2143	1897	2352	2341
Pediatrics							
<i>r</i>	0.49	0.51	0.49
<i>n</i>	2291	2313	2556
Family medicine							
<i>r</i>	0.52	0.50
<i>n</i>	2306	2283
Internal medicine							
<i>r</i>	0.54
<i>n</i>	2515
Surgery							
<i>r</i>
<i>n</i>

^a All Pearson product-moment correlations were statistically significant at $P < .001$.

^b A "high-stakes" examination was defined as a Comprehensive Osteopathic Medical Achievement Test (COMAT) in which the score was used for clerkship and clinical rotation evaluations. "Low-stakes" examination scores were not used for such purposes.

Abbreviation: OPP, osteopathic principles and practice.

In general, for 2011, high-stakes scores had higher correlations across subjects than low-stakes scores. However, COMAT OPP scores had lower correlations with COMLEX-USA Level 2-CE scores in the 2011 high-stakes group ($r=0.38$) than in the 2011 low-stakes group ($r=0.45$), probably because of a substantial difference in sample size, from 1556 to 749.

Table 4 provides results from regression analyses with scores from 7 COMAT subject examinations used to predict COMLEX-USA Level 2-CE scores for both low- and high-stakes groups in 2011 and the high-stakes group in 2012. All 3 regression analyses showed scores from internal medicine as the strongest predictors of Level 2-CE performance, whereas OPP scores were not a significant predictor

Table 3.
Correlations^a Between COMAT and COMLEX-USA Level 2-CE
Scores of Osteopathic Medical Students in 2011 and 2012

Group ^b	Obstetrics- Gynecology	OPP	Psychiatry	Pediatrics	Family Medicine	Internal Medicine	Surgery
2011							
Low stakes							
<i>r</i>	0.52	0.45	0.51	0.55	0.53	0.62	0.55
<i>n</i>	980	1556	532	866	784	529	872
High stakes							
<i>r</i>	0.58	0.38	0.52	0.57	0.61	0.64	0.61
<i>n</i>	1342	749	1338	1262	1504	1495	1478
2012							
High stakes							
<i>r</i>	0.60	0.45	0.50	0.59	0.58	0.65	0.61
<i>n</i>	2815	1879	2351	2566	2538	2776	2787

^a All Pearson product-moment correlations were statistically significant at $P < .001$.

^b A "high-stakes" examination was defined as a Comprehensive Osteopathic Medical Achievement Test (COMAT) in which the score was used for clerkship and clinical rotation evaluations. "Low-stakes" examination scores were not used for such purposes.

Abbreviations: COMLEX-USA Level 2-CE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Cognitive Evaluation; OPP, osteopathic principles and practice.

in 2011 but became significant in 2012. Both years had similar amounts of variance explained by 7 COMAT subject examinations for the high-stakes groups (adjusted $R^2=0.63$ for both 2011 and 2012), substantially more than for the 2011 low-stakes group (adjusted $R^2=0.47$).

Discussion

In the current study, we investigated the relationship between the COMAT subject examination scores and the COMLEX-USA Level 2-CE scores through correlational and regression analyses. We used data from the 2011 and 2012 COMAT cycles, focusing on students with scores in the high-stakes category whose schools used COMAT subject examinations as part of clerkship evaluations.

The results confirmed our hypotheses that intercorrelations between COMAT subject examination scores

were lower than the correlations between those scores and scores for the comprehensive COMLEX-USA Level 2-CE. One reason for the lower intercorrelations might be the number of items in the COMAT subject examinations, which is a quarter of the number of items in the COMLEX-USA series, which could affect the internal consistency of the shorter examinations. The intercorrelations we found were similar to those found for comparable allopathic subject and licensure examination scores in a smaller sample.¹² Transferable skills may be important contributors to performance, as suggested elsewhere.¹¹ In light of the test blueprints, one would expect correlations between COMAT and COMLEX-USA Level 2-CE scores to be higher than correlations between any 2 distinct subject areas, because COMLEX-USA comprises several disciplines, some of which overlap with the COMAT subject examinations.

Table 4.
Regression Models With Scores From COMAT Predicting COMLEX-USA Level 2-CE
Scores of Osteopathic Medical Students in 2011 and 2012

Predictors	2011				2012	
	Low Stakes ^a		High Stakes ^a		High Stakes ^a	
	Standardized Coefficients	P Value	Standardized Coefficients	P Value	Standardized Coefficients	P Value
Obstetrics-gynecology	0.71	.48	5.85	<.001	7.36	<.001
OPP	0.58	.57	1.66	.097	9.25	<.001
Psychiatry	1.74	.083	5.21	<.001	4.68	<.001
Pediatrics	2.83	.005	4.93	<.001	8.77	<.001
Family medicine	1.92	.055	7.07	<.001	7.11	<.001
Internal medicine	7.24	<.001	7.21	<.001	12.86	<.001
Surgery	3.64	<.001	6.42	<.001	9.21	<.001
Adjusted R²	.47		.63		.63	

^a A "high-stakes" examination was defined as a Comprehensive Osteopathic Medical Achievement Test (COMAT) in which the score was used for clerkship and clinical rotation evaluations. "Low-stakes" examination scores were not used for such purposes.

Abbreviations: COMLEX-USA Level 2-CE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-CognitiveEvaluation; OPP, osteopathic principles and practice.

All samples showed moderate to high correlation between COMAT scores and Level 2-CE scores. Generally, primary care subjects, such as internal medicine and family medicine, had higher correlations with Level 2-CE scores than non-primary care subjects, such as psychiatry. This finding may arise from the fact that although all the test blueprints are osteopathically oriented, the content domain for psychiatry, for example, differs distinctly from those for the other COMAT subject examinations or COMLEX-USA Level 2-CE. Furthermore, to our knowledge, some schools have reported using the COMAT OPP examination either as a comprehensive examination for their OPP curriculum at the end of the second year or as a third-year end-of-clerkship examination. This variable use of COMAT OPP scores probably contributes to the lowest observed intercorrelations of the 7 subjects. In contrast, COMAT surgery scores showed higher correlation with Level 2-CE scores than did COMAT family medicine scores.

A comparable study investigated the relationship between scores on the National Board of Medical Examiners (NBME) subject examinations and the United States Medical Licensing Examination (USMLE) Step 1 and Step 2 Clinical Knowledge.¹² The researchers found that the NBME surgery examination scores had a higher correlation with USMLE Step 2 Clinical Knowledge scores than did the NBME family medicine scores but a lower correlation than the NBME internal medicine scores. Overall, the correlational results from the USMLE study¹² are similar to those in the current study.

Arguably, when students take COMAT for high-stakes purposes, as for use in grading and determining pass-fail status for clerkships and clinical rotations, their COMAT performance will probably better reflect their actual ability and skills. Subsequently, COMAT scores in a high-stakes group would probably better predict COMLEX-USA Level 2-CE scores. The comparison of results between high-stakes and low-stakes groups

strongly support this supposition, especially for the 2011 sample, which included a sizeable proportion of students (47.0%) who were not in a high-stakes scenario. When we removed the low-stakes scores from the analysis, the results indicated generally higher correlations with Level 2-CE scores and explained more of the variance in Level 2-CE scores. These findings suggest that schools using the COMAT subject examinations for high-stakes purposes can better predict COMLEX-USA Level 2-CE performance.

Even though we found that COMAT scores had moderate-to-high correlations with COMLEX-USA Level 2-CE scores and that COMAT performance explained a substantial amount of variance in Level 2-CE performance, 37% of the variance was still unexplained by the current models. As a future study, a multilevel analysis with students nested in schools may be considered, wherein variation within schools (eg, rotation length, rotation timing, and various uses of COMAT scores) can be further analyzed. To determine why the lowest correlations were between COMLEX-USA Level 2-CE and COMAT OPP scores, a follow-up study investigating regional differences in OPP curricula and possible effects on item-level performances is forthcoming.

Conclusion

The results of the current study show that, as we hypothesized, COMAT subject examination scores were moderately intercorrelated, with higher correlations between the COMAT and the COMLEX-USA Level 2-CE scores. Further, COMAT performance was predictive of COMLEX-USA Level 2-CE performance.

Author Contributions

Drs Li and Song provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; Dr Kalinowski drafted the article or revised it critically for important intellectual content; and Dr Bates gave final approval of the version of the article to be published.

References

1. American Association of Colleges of Osteopathic Medicine. *Osteopathic Medical College Information Book*. Chevy Chase, MD: American Association of Colleges of Osteopathic Medicine; 2013.
2. Cavalieri TA, Shen L, Slick GL. Predictive validity of osteopathic medical licensing examinations for osteopathic medical knowledge measured by graduate written examinations. *J Am Osteopath Assoc*. 2003;103(7):337-342.
3. Langenau E, Pugliano G, Roberts W. Competency-based classification of COMLEX-USA cognitive examination test items. *J Am Osteopath Assoc*. 2011;111(6):396-402. <http://www.jaoa.org/content/111/6/396>. Accessed September 30, 2013.
4. Meoli FG, Wallace WS, Kaiser-Smith J, Shen L. Relationship of osteopathic medical licensure examinations with undergraduate admission measures and predictive value of identifying future performance in osteopathic principles and practice/osteopathic manipulative medicine courses and rotations. *J Am Osteopath Assoc*. 2002;102(11):615-620. <http://www.jaoa.org/content/102/11/615>. Accessed September 30, 2013.
5. Roberts WL, Solomon M, Langenau E. An investigation of construct validity of humanistic clinical skills on a medical licensure examination. *Patient Educ Couns*. 2011;82(2):214-221. doi:10.1016/j.pec.2010.03.016.
6. Sevensma SC, Navarre G, Richards RK. COMLEX-USA and in-service examination scores: tools for evaluating medical knowledge among residents. *J Am Osteopath Assoc*. 2008;108(12):713-716. <http://www.jaoa.org/content/108/12/713>. Accessed September 30, 2013.
7. Wright BD, Bezruczko N. *Evaluation of COMLEX-USA*. Chicago, IL: MESA Psychometric Laboratory; 1999.
8. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving*. Cambridge, MA: Harvard University Press; 1978.
9. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ*. 2006;40(7):618-623. doi:10.1111/j.1365-2929.2006.02511.x.
10. Norcini J, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ*. 2002;36(9):853-859. doi:10.1046/j.1365-2923.2002.01293.x.
11. Mattick K, Dennis I, Bradley P, Bligh J. Content specificity: is it the full story? statistical modelling of a clinical skills examination. *Med Educ*. 2008;42(6):589-599. doi:10.1111/j.1365-2923.2008.03020.x.
12. Zahn CM, Saguil A, Artino AR Jr, et al. Correlation of National Board of Medical Examiners scores with United States Medical Licensing Examination Step 1 and Step 2 scores. *Acad Med*. 2012;87(10):1348-1354. doi:10.1097/ACM.0b013e31826a13bd.

© 2014 American Osteopathic Association