A Validity Study of COMLEX-USA Level 2-CE and COMAT Clinical Subjects: Concurrent and Predictive Evidence

Kimberly M. Hudson, PhD
Tsung-Hsun Tsai, PhD
Charles Finch, DO
Joel L. Dickerman, DO
Silu Liu, MPH
Ling Shen, MS

ABSTRACT

Background The Comprehensive Osteopathic Medical Licensure Examination (COMLEX-USA) Level 2–Cognitive Examination (CE) and the Comprehensive Osteopathic Medical Achievement Test (COMAT) are administered to similar populations (third- and fourth-year osteopathic students) at similar points in time. Examining the relationship between scores on the 2 assessments that measure similar constructs ultimately supports the validity of both.

Objective The purpose of this study is to provide empirical evidence of the concurrent and predictive validity of COMAT and COMLEX-USA Level 2-CE.

Methods In 2018, first-attempt scores on Level 2-CE were aggregated from June 2015 to May 2018 and matched with first-attempt scores on each COMAT clinical subject. We conducted correlational analyses between performance on COMAT and Level 2-CE, and COMAT scores and Level 2-CE discipline subscores. Additionally, we used multivariate regression to analyze the predictive relationship between performance on all COMAT clinical subjects and Level 2-CE.

Results The results from correlational analyses indicated statistically significant, positive associations between COMAT and Level 2-CE scores ($r = 0.49–0.68$, $P < .0001$), and statistically significant, but slightly weaker relationships between COMAT scores and Level 2-CE discipline subscores ($r = 0.31–0.60$, $P < .0001$). Furthermore, results from the multiple regression indicated that scores on COMAT explained 68% of the variance in Level 2-CE scores, and that COMAT internal medicine and emergency medicine were weighted more heavily than other specialties.

Conclusions The findings from this study can inform assessment practices by supporting the use of COMAT for osteopathic medical schools that do not administer COMAT.

Introduction

Students enrolled in osteopathic medical schools are required to pass a series of licensing examinations, the Comprehensive Medical Licensing Examination (COMLEX-USA), in order to practice osteopathic medicine in unsupervised clinical settings. The Comprehensive Osteopathic Medical Achievement Test (COMAT) is typically administered to osteopathic students after clinical rotations or clerkships to assess knowledge with respect to core disciplines. Although these 2 assessment programs have different purposes, they both measure similar knowledge (ie, clinical competencies). Empirical evidence of the relationship between performance on both examinations supports the validity of these assessment programs and is a critical component in educational measurement.[1] Now that students attending osteopathic and allopathic medical schools can match to the same residency program under the Accreditation Council for Graduate Medical Education's single accreditation system, it is imperative for residency program directors and graduate medical educators to have sufficient information about COMLEX-USA and COMAT in order to make informed decisions about applicants graduating from osteopathic medical schools. Performance on COMAT and its correlation with performance on COMLEX-USA may become increasingly important in discussions regarding moving licensure examinations to pass/fail.

The 4-part COMLEX-USA series includes Level 1, Level 2–Cognitive Evaluation (CE), Level 2–Performance Evaluation (PE), and Level 3. Prior to 2018, all levels were designed to measure proficiency across the same 9 patient presentations and 6 physician tasks. A new test blueprint was implemented in 2018 for Level 3, and in 2019 for Levels 1, 2-CE, and 2-PE. Levels 1, 2-CE, and 3 are computer-based examinations, while Level 2-PE is performance-based. Level 2-CE covers clinical concepts required for medical problem-solving with a concentration on collecting a thorough patient history, analyzing physical examination findings, and making appropriate medical diagnoses.[2] COMAT is a series of 8 computer-based examinations, each focusing on a specific discipline.[3] Some

DOI: http://dx.doi.org/10.4300/JGME-D-19-00157.1

osteopathic medical schools use COMAT scores as an evaluation of students' clerkship or clinical rotation.[4]

Despite having different purposes, COMAT and Level 2-CE have several similarities. First, physicians write items and review content employing the same item writing and test development principles. Secondly, each COMAT represents a unique medical discipline that is also represented on Level 2-CE. Moreover, both examination programs are administered in a time-measured environment and consist of multiple-choice questions. Lastly, osteopathic students typically take COMAT and Level 2-CE during their third or fourth year, although no widespread rule is applied to the sequence.

Several research studies have investigated the concurrent and predictive validity of assessments that purport to measure similar constructs. Prior research has supported the relationship between students' outcomes on clinical rotations and medical licensure examinations.[4,5] After investigating the correlation between osteopathic school performance and Level 2-CE, researchers found a stronger relationship between second-year osteopathic school performance and Level 2-CE than first-year performance ($r = 0.70$–0.59; $N = 171$, $N = 86$, respectively).[5] Two years following the launch of COMAT, Li and colleagues found moderate correlations between performance on each of the 7 COMAT clinical subjects and Level 2-CE. For most clinical subjects, the correlations were stronger for the second year of administration compared to the first year of administration ($r = 0.45$–0.65, $r = 0.38$–0.64, respectively).[4] These findings were based on a relatively small dataset shortly after the launch of COMAT.

Similar studies have been conducted on the National Board of Medical Examiners' clinical subject "shelf" examinations.[6–9] Zahn and colleagues found performance on shelf examinations accounted for 60% of the variance in Step 2–Clinical Knowledge (CK), where primary care specialties were weighted more heavily ($r = 0.64$, $N = 507$).

The current study aimed to answer the following research questions: (1) What is the correlation between performance on Level 2-CE and performance on COMAT clinical subjects? (2) What is the correlation between Level 2-CE discipline subscores and COMAT clinical subject scores? (3) Does performance on each COMAT clinical subject predict performance on Level 2-CE?

What was known and gap
The COMLEX-USA Level 2–CE and the COMAT are administered to osteopathic students at similar points in time. Under the Accreditation Council for Graduate Medical Education's single accreditation system, residency program directors need sufficient information about COMLEX-USA and COMAT in order to make informed decisions about applicants graduating from osteopathic medical schools.

What is new
A validity study to analyze the predictive relationship between performance on all COMAT clinical subjects and Level 2-CE.

Limitations
Study included only students with complete data (eg, graduation year) and those enrolled in osteopathic medical schools that use COMAT for evaluative purposes. The multivariate regression model was based on a smaller sample size consisting of students who took all 8 COMAT.

Bottom line
There is a strong positive relationship among performance on COMAT and Level 2-CE, with up to 68% of variance explained.

Methods

Participants

In order to address the research questions, first-attempt scores on Level 2-CE were aggregated from June 2015 to May 2018 ($N = 17\,991$). Data were then matched with first-attempt scores on each COMAT clinical subject using unique identifiers. Students with incomplete data (eg, missing graduation year) were excluded from the analyses. Only students who had taken COMAT prior to Level 2-CE were included in the study. Lastly, the dataset was limited to students who attended osteopathic medical schools that use COMAT scores for high-stakes evaluative purposes because administrative conditions are more similar to that of COMLEX-USA. Moreover, prior research has supported stronger relationships when COMAT is used for evaluative purposes.[4]

The study was approved as human subjects research through expedited review by the Institutional Review Board at the University of Illinois at Chicago.

Analyses

To address the first and second research questions, Pearson product-moment correlation coefficients were calculated between COMAT scores and Level 2-CE scores and between COMAT scores and Level 2-CE discipline subscores. Multivariate regression analysis was used to determine how much variance in Level 2-CE scores was explained by COMAT scores. A significance level of .01 was used to determine statistical significance unless otherwise specified. The Bonferroni correction was used to control for multiple comparisons and family-wise error rate, which statistically adjusts when students took more than 1 COMAT. All data management and analyses were completed in SAS 9.4 (SAS Institute Inc, Cary, NC).

TABLE 1
COMLEX–USA Level 2-CE and COMAT Performance by Cohort

	Cohort			Overall
	2016	2017	2018	
Student count (%)	3808 (29)	4174 (32)	5143 (39)	13 125 (100)
Level 2-CE mean (SD)	533.45 (100.77)	539.79 (101.53)	546.94 (106.11)	540.75 (103.27)
COMAT EM mean (SD)	94.71 (8.87)	98.09 (9.67)	99.25 (9.02)	98.74 (9.31)
COMAT FM mean (SD)	100.93 (10.48)	100.70 (9.20)	101.08 (9.69)	100.92 (9.78)
COMAT IM mean (SD)	101.44 (10.69)	99.66 (8.52)	101.02 (9.87)	100.71 (9.74)
COMAT OB mean (SD)	98.49 (11.29)	100.17 (9.15)	100.78 (9.22)	99.96 (9.85)
COMAT OP mean (SD)	101.62 (10.53)	101.50 (9.92)	100.92 (9.72)	101.30 (10.01)
COMAT PD mean (SD)	99.94 (10.46)	105.94 (7.79)	101.24 (9.61)	102.44 (9.63)
COMAT PS mean (SD)	99.59 (11.14)	101.12 (8.99)	102.11 (9.47)	101.08 (9.87)
COMAT SU mean (SD)	100.52 (11.01)	101.33 (9.56)	101.06 (9.50)	101.00 (9.95)

Abbreviations: COMLEX-USA, Comprehensive Osteopathic Medical Licensure Examination; Level 2-CE, Level 2–Cognitive Evaluation; COMAT, Comprehensive Osteopathic Medical Achievement Test; EM, emergency medicine; FM, family medicine; IM, internal medicine; OB, obstetrics and gynecology; OP, osteopathic principles and practice; PD, pediatrics; PS, psychiatry; SU, surgery.

Results

Descriptive Statistics

After applying the exclusion criteria, approximately 27% (N = 4866) of student records were removed. The majority of excluded records were because students did not attend an osteopathic medical school that administered COMAT as a high-stakes assessment prior to Level 2-CE. The 13 125 students used for analyses represented the 2016, 2017, and 2018 cohorts: 3808 (29%), 4174 (32%), and 5143 (39%), respectively. Approximately 56% (7350 of 13 125) of students took 7 or 8 COMAT clinical subjects. TABLE 1 shows the mean and standard deviation of scores on Level 2-CE and each COMAT clinical subject overall and by cohort. Although the overall mean Level 2-CE score was 540.75 (SD = 103.27), there was some variation in Level 2-CE performance across cohorts. The 2018 cohort performed the highest (M = 546.94, SD = 106.11) and the 2016 cohort performed the lowest (M = 533.45, SD = 100.77). The mean COMAT scores were stable across cohorts.

TABLE 2 shows the descriptive statistics by each COMAT. Due to the recent launch of emergency medicine (EM), the number of administrations was relatively lower than other clinical subjects. The mean EM score was lower than the mean score of other clinical subjects (M = 98.74, SD = 9.31). The mean pediatrics score was the highest (M = 102.44, SD = 9.63). Results from 1-way analysis of variance indicated that there were statistically significant differences among scores on each COMAT clinical subject ($F(7,79498) = 71.58$, $P < .0001$). However, findings indicated that most of the significant differences were due to differential performance on EM, obstetrics and gynecology, osteopathic principles and practice, surgery, and pediatrics.

Validity Evidence

Correlational analyses were conducted to examine the relationship between performance on COMAT and Level 2-CE. TABLE 3 shows the correlations between COMAT scores and Level 2-CE total scores, and between COMAT scores and Level 2-CE discipline subscores. A moderate to high correlation was observed between COMAT and Level 2-CE scores (r = 0.49–0.68, $P < .001$). Meaning, COMAT scores accounted for 24% to 46% of the variance in Level 2-CE scores when examined independently. Correlations between COMAT scores and their corresponding Level 2-CE discipline subscores was the greatest for internal medicine (IM; r = 0.60, $P < .0001$) and the lowest for psychiatry (r = 0.31, $P < .0001$). Overall, the correlations between scores on COMAT and the Level 2-CE were equal to or greater than the correlations between scores on COMAT and Level 2-CE discipline subscores for all clinical subjects.

A multivariate regression model was used to predict performance on Level 2-CE scores using data from 1143 students who took all 8 COMAT clinical subjects. For each COMAT subjects, TABLE 4 shows the parameter estimate (β), the standard error of the estimate (SE(β)), and the associated significance value. We confirmed that the model assumptions were met prior to analyses; specifically, the variance inflation factor value for parameter estimates was less than 2, indicating no multicollinearity, and the Q–Q plot showed normally distributed error terms.

The multivariate regression model was statistically significant, where performance on each COMAT contributed to performance on Level 2-CE ($F(8,1134) = 299.08$, $P < .0001$). COMAT performance across all 8 subjects accounted for 68% of the variation in Level 2-CE performance. Furthermore,

TABLE 2
Descriptive Statistics by COMAT Clinical Subjects (N = 13 125)

COMAT Subject	Students	COMAT, Mean (SD)	Level 2-CE, Mean (SD)	Level 2-CE Passing Rate
EM	1983	98.74 (9.31)	541.47 (105.70)	91.63%
FM	10 547	100.92 (9.78)	537.86 (102.95)	91.78%
IM	11 732	100.71 (9.74)	539.70 (103.41)	91.92%
OB	11 510	99.96 (9.85)	540.18 (103.74)	91.86%
OP	9991	101.30 (10.01)	541.49 (102.85)	92.57%
PD	11 349	102.44 (9.63)	539.41 (103.65)	91.77%
PS	11 009	101.08 (9.87)	539.75 (103.65)	91.75%
SU	11 385	101.00 (9.95)	540.26 (103.92)	91.86%

Abbreviations: COMAT, Comprehensive Osteopathic Medical Achievement Test; Level 2-CE, Level 2–Cognitive Evaluation; EM, emergency medicine; FM, family medicine; IM, internal medicine; OB, obstetrics and gynecology; OP, osteopathic principles and practice; PD, pediatrics; PS, psychiatry; SU, surgery.

IM and EM had more impact on the Level 2-CE performance than other COMAT subjects ($\beta = 2.27$ and $\beta = 2.20$, respectively). Due to the relatively lower volume of students who had taken COMAT EM, we also analyzed data excluding COMAT EM and found comparable results ($F(7,6833) = 1997.26$, $P < .0001$). However, slightly less variance was explained by performance on 7 subjects than on 8 subjects (adjusted $R^2 = 0.67$ versus adjusted $R^2 = 0.68$, respectively).

Discussion

When COMAT is administered in part to evaluate students' clinical rotations, scores on COMAT explain 68% of the variance in Level 2-CE scores. In addition, there are significant correlations among scores on COMAT and Level 2-CE discipline subscores. This research expands on prior studies that used smaller sample sizes and did not include COMAT EM.[4]

The evidence presented suggests a moderate, statistically significant relationship between scores on COMAT and Level 2-CE. Similar to findings from prior research that compared NBME "shelf" examinations and Step 2-CK,[6] the strongest relationships were evidenced between COMAT IM and Level 2-CE scores, and COMAT EM and Level 2-CE scores ($r = 0.68$, $P < .0001$ for both examinations). The relationship was further exemplified by the moderate, statistically significant correlations between COMAT scores and their corresponding Level 2-CE discipline subscores ($r = 0.31$–0.61, $P < .0001$). Overall, there were stronger relationships between primary care subscores (family medicine and IM), likely due to the emphasis on Level 2-CE as a generalist examination. These results are consistent with prior literature indicating stronger correlations between primary care specialties.[6] These findings support the concurrent validity between COMAT and Level 2-CE.

TABLE 3
Correlations Between COMAT Score and Level 2-CE Score and Level 2-CE Discipline Subscores

COMAT Subject	Student Count	Correlation With Level 2-CE Score[a]	Correlation With Level 2-CE Discipline Subscores[a]
EM	1983	0.68	0.53
FM	10 547	0.64	0.53
IM	11 732	0.68	0.60
OB	11 510	0.63	0.44
OP	9991	0.49	0.49
PD	11 349	0.61	0.50
PS	11 009	0.58	0.31
SU	11 385	0.66	0.39

Abbreviations: COMAT, Comprehensive Osteopathic Medical Achievement Test; Level 2-CE, Level 2–Cognitive Evaluation; EM, emergency medicine; FM, family medicine; IM, internal medicine; OB, obstetrics and gynecology; OP, osteopathic principles and practice; PD, pediatrics; PS, psychiatry; SU, surgery.
[a] All correlation coefficients were statistically significant ($P < .0001$).

TABLE 4
Multivariate Regression Model Predicting Level 2-CE Score Using 8 COMAT Scores

Predictor	β	SE (β)	*P* Value
EM	2.20	0.27	< .0001
FM	1.59	0.26	< .0001
IM	2.27	0.26	< .0001
OB	1.31	0.27	< .0001
OP	1.27	0.21	< .0001
PD	1.37	0.24	< .0001
PS	1.04	0.23	< .0001
SU	1.21	0.26	< .0001
Adjusted R^2	0.68		

Abbreviations: Level 2-CE, Level 2–Cognitive Evaluation; COMAT, Comprehensive Osteopathic Medical Achievement Test; EM, emergency medicine; FM, family medicine; IM, internal medicine; OB, obstetrics and gynecology; OP, osteopathic principles and practice; PD, pediatrics; PS, psychiatry; SU, surgery.

The results from the multivariate linear regression provide strong evidence that performance on COMAT predicts performance on Level 2-CE, supporting the predictive validity of examination programs that are designed to measure similar constructs. High performance on COMAT IM and EM is the strongest predictor of high performance on Level 2-CE because of the emphasis of IM and EM on Level 2-CE. Another factor that may contribute to the greater weight of EM scores than other subjects may be that COMAT EM is typically the last COMAT subject taken by students.

The ongoing move to joint accreditation makes this study an important component of the broader graduate medical education context. Now that students attending osteopathic and allopathic medical schools participate in the Match, the findings of this research can inform residency program directors when making decisions about osteopathic applicants.

Practical considerations should be taken into account when interpreting the results of this research. First, we included only students with complete data (eg, graduation year) and those who were enrolled in osteopathic medical schools that use COMAT for evaluative purposes. Therefore, these results may not hold for the entire population. Second, the multivariate regression model was based on a smaller sample size consisting of students who took all 8 COMAT due to the relative lower volume of EM administrations. Additionally, this research did not explore any other factors that may be related to Level 2-CE scores (eg, Level 1 scores, clerkship grades, sequencing of COMAT administrations). Furthermore, although COMAT and COMLEX-USA are designed to measure similar topics (eg, disciplines) and have similar computer-based formats, the purposes of the assessments differ. While COMAT is designed to measure discipline-specific knowledge in order for osteopathic medical schools to assess student competencies, Level 2-CE emphasizes primary care specialties as a generalist examination; therefore, discipline subscores on Level 2-CE are less reliable than total scores.

Future research can address these areas by including additional variables (eg, clerkship grades, scores on Levels 1 and 2-PE) or by examining the timing of COMAT and Level 2-CE as other factors that may influence performance on either examination.

Conclusions

The findings from this study indicate that there is a strong positive relationship among performance on COMAT and Level 2-CE, with up to 68% of variance explained. This study is applicable to osteopathic medical schools that do not administer COMAT for evaluative purposes as it indicates a moderate significant relationship between 2 assessments administered to osteopathic students. These findings should be taken into consideration when evaluating osteopathic applicants during the residency application process.

References

1. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
2. National Board of Osteopathic Medical Examiners. COMLEX-USA Bulletin of Information 2018–2019. https://www.nbome.org/Content/Exams/COMLEX-USA/COMLEX-USA_Bulletin_of_Information.pdf. Accessed September 4, 2019.
3. National Board of Osteopathic Medical Examiners. Comprehensive Osteopathic Medical Achievement Test: Test Administration Guide 2018–2019. https://www.nbome.org/wordpress/wp-content/uploads/2018/06/COMAT-Administration-Guide-2018-19.pdf. Accessed September 4, 2019.
4. Li F, Kalinowski KE, Song H, Bates BP. Relationships between the Comprehensive Medical Achievement Test (COMAT) subject examinations and the COMLEX-USA Level 2–Cognitive Evaluation. *J Am Osteopath Assoc*. 2014;114(9):714–721. doi:10.7556/jaoa.2014.140.
5. Hartman SE, Bates BP, Sprafka SA. Correlation of scores for the Comprehensive Osteopathic Medical Licensing Examination with osteopathic medical school grades. *J Am Osteopath Assoc*. 2001;101(6);347–349.
6. Zahn CM, Saguil A, Artino AR Jr, Dong T, Ming G, Servey JT, 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.
7. Dong T, Copeland A, Gangidine M, Schreiber-Gregory D, Ritter EM, Durning SJ. Factors associated with surgery clerkship performance and subsequent USMLE step scores. *J Surg Educ*. 2018;75(5):1200–1205. doi:10.1016/j.surg.2018.02.17.
8. Myles T, Galvez-Myles R. USMLE Step 1 and 2 scores correlate with family medicine clinical and examination scores. *Fam Med*. 2003;35(7):510–513.
9. Ogunyemi D, De Taylor-Harris S. NBME obstetrics and gynecology clerkship final examination scores: predictive value of standardized tests and demographic factors. *J Reprod Med*. 2004;49(12):978–982.

Kimberly M. Hudson, PhD, is Manager, Psychometrics and Research, National Board of Osteopathic Medical Examiners;

Tsung-Hsun Tsai, PhD, is Associate Vice President, Assessment Services and Research, National Board of Osteopathic Medical Examiners; **Charles Finch, DO,** is Chair, Department of Integrated Medicine, and Professor, Department of Emergency Medicine, Arizona College of Osteopathic Medicine and Midwestern University; **Joel L. Dickerman, DO,** is Vice President, Assessment Services, National Board of Osteopathic Medical Examiners; **Silu Liu, MPH,** is Senior Research Associate, National Board of Osteopathic Medical Examiners; and **Ling Shen, MS,** is Research Associate, National Board of Osteopathic Medical Examiners.

Funding: The authors report no external funding source for this study.

Conflict of interest: Five of the 6 authors were employed by the National Board of Osteopathic Medical Examiners at the time this study was conducted.

The authors would like to thank the advisory committee for COMLEX-USA Level 2-CE, who reviewed and provided helpful suggestions on the research. We would also like to thank the National Board of Osteopathic Medical Examiners staff, specifically Dr. John Gimpel, Karen Huelsman, and Shirley Bodett, who provided comments on prior versions of this manuscript.

Corresponding author: Kimberly M. Hudson, PhD, National Board of Osteopathic Medical Examiners, 8765 W Higgins Road, Suite 200, Chicago, IL 60631, 773.714.0622 ext. 266, kimberly.shay86@gmail.com

Received February 28, 2019; revision received August 11, 2019; accepted August 12, 2019.