Relationship Between COMLEX-USA Scores and Performance on the American Osteopathic Board of Emergency Medicine Part I Certifying Examination

Feiming Li, PhD; John R. Gimpel, DO; Ethan Arenson, MA; Hao Song, PhD; Bruce P. Bates, DO; and Fredric Ludwin, DO

From the National Board of Osteopathic Medical Examiners (NBOME) (Drs Li, Gimpel, Song, and Bates and Mr Arenson) and from the American Osteopathic Board of Emergency Medicine (AOBEM) (Dr Ludwin).

Financial Disclosures: Dr Li is a psychometrician at the NBOME; Dr Gimpel is the president and chief executive officer of the NBOME; Mr Arenson is a research associate at the NBOME; Dr Song is the senior director of psychometrics and research at the NBOME; and Dr Bates is the senior vice president for cognitive testing at the NBOME. Dr Ludwin is chairman of the AOBEM.

Address correspondence to Feiming Li, PhD, National Board of Osteopathic Medical Examiners, 8765 W Higgins Rd, Suite 200, Chicago, IL 60631-4174.

E-mail: fli@nbome.org

Submitted October 1, 2013; revision received January 3, 2014; accepted February 26, 2014.

Context: Few studies have investigated how well scores from the Comprehensive Osteopathic Medical Licensing Examination-USA (COMLEX-USA) series predict resident outcomes, such as performance on board certification examinations.

Objectives: To determine how well COMLEX-USA predicts performance on the American Osteopathic Board of Emergency Medicine (AOBEM) Part I certification examination.

Methods: The target study population was first-time examinees who took AOBEM Part I in 2011 and 2012 with matched performances on COMLEX-USA Level 1, Level 2-Cognitive Evaluation (CE), and Level 3. Pearson correlations were computed between AOBEM Part I first-attempt scores and COMLEX-USA performances to measure the association between these examinations. Stepwise linear regression analysis was conducted to predict AOBEM Part I scores by the 3 COMLEX-USA scores. An independent t test was conducted to compare mean COMLEX-USA performances between candidates who passed and who failed AOBEM Part I, and a stepwise logistic regression analysis was used to predict the log-odds of passing AOBEM Part I on the basis of COMLEX-USA scores.

Results: Scores from AOBEM Part I had the highest correlation with COMLEX-USA Level 3 scores (.57) and slightly lower correlation with COMLEX-USA Level 2-CE scores (.53). The lowest correlation was between AOBEM Part I and COMLEX-USA Level 1 scores (.47). According to the stepwise regression model, COMLEX-USA Level 1 and Level 2-CE scores, which residency programs often use as selection criteria, together explained 30% of variance in AOBEM Part I scores. Adding Level 3 scores explained 37% of variance. The independent t test indicated that the 397 examinees passing AOBEM Part I performed significantly better than the 54 examinees failing AOBEM Part I in all 3 COMLEX-USA levels ($P<.001$ for all 3 levels). The logistic regression model showed that COMLEX-USA Level 1 and Level 3 scores predicted the log-odds of passing AOBEM Part I ($P=.03$ and $P<.001$, respectively).

Conclusion: The present study empirically supported the predictive and discriminant validities of the COMLEX-USA series in relation to the AOBEM Part I certification examination. Although residency programs may use COMLEX-USA Level 1 and Level 2-CE scores as partial criteria in selecting residents, Level 3 scores, though typically not available at the time of application, are actually the most statistically related to performances on AOBEM Part I.

Predictive studies of examinations are important because these studies help establish the validity of the examination against a specific criterion.[1,2] The problem with such studies, particularly in licensure and certification examinations, is that it can be difficult to measure the ideal criterion: professional competence.[2,3] In the context of medical licensure examinations, the validity of the examination is built through careful construction of content on the basis of practice analysis of, on one hand, practicing physicians and procedural rigorousness of examination analysis, and on the other hand, associating the licensed physician with clinical performance outcomes. Among many measures of professional performance, board certification pass-fail statuses are often used as a proxy measure for professional competence.[4-6]

The Comprehensive Osteopathic Medical Licensing Examination-USA (COMLEX-USA) is a 3-level licensing examination series for osteopathic physicians. In addition, residency program directors often use 2 of the COMLEX-USA scores (Level 1 and Level 2-Cognitive Evaluation [CE]) as partial criteria to select applicants for residency programs.[5-7] Few published studies have investigated how well COMLEX-USA scores predict resident outcomes, such as performance on board certification examinations.[8]

The American Osteopathic Board of Emergency Medicine (AOBEM) requires candidates to pass 3 examinations as a requirement for board certification. Part I is a computer-based multiple-choice examination. Part II is an oral examination that consists of clinical presentations involving either specific data or simulated patient encounters related to emergency medicine cases. In Part III, candidates submit 20 deidentified medical records of clinical emergency department patients. Eight of these patients must have been admitted to the hospital or transferred to another health care facility.

In the present study, we investigated the relationship between osteopathic medical students' performances on the COMLEX-USA series and their performances on the AOBEM Part I certification examination. For COMLEX-USA, a 3-digit standard score of 400 on Level 1 or Level 2-CE and a 3-digit standard score of 350 on Level 3 are required to pass the examination. For AOBEM Part I, examinees are required to earn a score of 500 to pass the examination. The research questions posited were the following: Can COMLEX-USA performance predict osteopathic emergency physicians' performance on AOBEM Part I? If yes, how well does every level of the COMLEX-USA series predict performance individually and collectively?

Methods

In the present study, we targeted first-time examinees who took AOBEM Part I in 2011 and 2012 and matched their first-time performances on COMLEX-USA Level 1, Level 2-CE, and Level 3 based on first name, last name, graduate year, and graduate school. Typically, the examinees who took AOBEM Part I took COMLEX-USA Level 1 seven years earlier, COMLEX-USA Level 2-CE five years earlier, and COMLEX-USA Level 3 three or four years earlier. The matched examinees who took COMLEX-USA earlier than 2002 were excluded because the data before 2002 were sparse (fewer than 15 candidates per year).

We performed an analysis of the Pearson correlations on first-attempt performance on AOBEM Part I and COMLEX-USA to measure the association between COMLEX-USA and AOBEM Part I performances. In addition, we performed a stepwise linear regression analysis to predict AOBEM Part I scores by the 3 COMLEX-USA scores. Lastly, we conducted independent t tests in which we compared mean COMLEX-USA performances between candidates who passed and candidates who failed AOBEM Part I, as well as a stepwise logistic regression to predict the log-odds of passing AOBEM Part I on the basis of COMLEX-USA scores. All of the analyses were conducted at α=.05 significance level.

Results

Of 560 AOBEM Part I examinees in 2011 and 2012, a total of 451 had all 3 COMLEX-USA scores obtained after 2002. This data set was used for all analyses. *Table 1* presents the descriptive statistics and passing rates for all 4 examinations. *Table 1* also provides the correlation between AOBEM Part I scores with scores from Levels 1, 2-CE, and 3 of COMLEX-USA. Scores from AOBEM Part I had the highest correlation with COMLEX-USA Level 3 scores (.57) and moderate correlation with COMLEX-USA Level 2-CE scores (.53). There was also a moderate, but slightly lower, correlation between AOBEM Part I scores and COMLEX-USA Level 1 scores (.47).

Table 2 presents the results of stepwise linear regression analysis. In step 1 of the analysis, we entered COMLEX-USA Level 1 and 2-CE scores, which may have factored into the selection criteria for residency programs. Together, the results of these 2 examinations explained 30% of variance in AOBEM Part I scores. In step 2, we entered COMLEX-USA Level 3 scores. Level 3 scores are generally not included as selection criteria as part of the residency program application because candidates generally take this examination during their resident training. Consequently, COMLEX-USA Level 3 scores were, in terms of time, closest to scores from AOBEM Part I. Level 3 scores added 7% additional variance to the model. In the final model, the 3 COMLEX-USA levels were all statistically significant predictors and explained 37% of variance of AOBEM Part I scores in total. According to the standardized coefficients, COMLEX-USA Level 3 scores were the strongest predictor, which was consistent with the fact that those scores had the highest correlation with AOBEM Part I performance.

We conducted independent *t* tests to compare the performance differences in each of the 3 COMLEX-USA levels, between the examinees who failed AOBEM Part I and the examinees who passed AOBEM Part I. *Table 3* shows that the 397 examinees who passed AOBEM Part I performed significantly better than the 54 examinees who failed AOBEM Part I in all 3 COMLEX-USA levels (P<.001): 63 points higher in Level 1 scores, 75 points higher in Level 2-CE scores, and 116 points higher in Level 3 scores. These differences indicate that the mean scores for the passing group were higher than those for the failing group by roughly 1 standard deviation for each COMLEX-USA level.

Table 4 provides the results from stepwise logistic regression. Similar to the results in *Table 3*, in step 1 we entered COMLEX-USA Level 1 and Level 2-CE scores as predictors. For both examinations, higher scores significantly increased the log-odds of passing AOBEM Part I. In step 2, after we entered COMLEX-USA Level 3 scores into the model, Level 1 and Level 3 scores significantly predicted the log-odds of passing AOBEM Part I (P=.03 and P<.001, respectively), but Level 2-CE scores were no longer statistically significant (P=.06). This result is most likely attributable to the high correlation between Level 2-CE and Level 3 scores. When 2 highly correlated predictors (Level 1 and Level 3 scores) explain a dependent variable, a relatively weaker predictor (Level 2-CE) may lose statistical significance and hence be dropped from the model.

To better illustrate how the increase in COMLEX-USA scores could improve the probability (rather than the log-odds) of passing AOBEM Part I, the plots for the probability of passing AOBEM vs each COMLEX-USA level are provided in the *Figure*. In the first portion of the *Figure*, for example, examinees with Level 1

Table 1.
COMLEX-USA and AOBEM Certification Examination Performance and Correlation (N=451)

Examination	Score, mean (SD)	Passing, %	Correlation With AOBEM Part I
AOBEM Part I	619.2 (101.2)	88.0	NA
COMLEX-USA Level 1	476.7 (70.3)	87.1	0.47
COMLEX-USA Level 2-CE	485.7 (78.6)	84.9	0.53
COMLEX-USA Level 3	496.5 (118.0)	90.9	0.57

Abbreviations: AOBEM, American Osteopathic Board of Emergency Medicine; CE, Cognitive Evaluation; COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA; NA, not applicable; SD, standard deviation.

Table 2.
Stepwise Linear Regression Model for AOBEM Part I Performance
as Predicted by COMLEX-USA Performance (N=451)

Examination	Unstandardized Coefficients	Standard Error	Standardized Coefficients	P Value	ΔR^2
Regression Step 1					.30
COMLEX-USA Level 1	0.30	0.08	3.79	<.001	
COMLEX-USA Level 2-CE	0.50	0.07	7.19	<.001	
Regression Step 2					.07
COMLEX-USA Level 1	0.16	0.08	2.13	.033	
COMLEX-USA Level 2-CE	0.27	0.07	3.61	<.001	
COMLEX-USA Level 3	0.31	0.04	7.08	<.001	
Total Model Adjusted R^2					.37

Abbreviations: AOBEM, American Osteopathic Board of Emergency Medicine; CE, Cognitive Evaluation; COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA; ΔR^2, change in multivariate coefficient of determination.

scores of 400 had a predicted probability of .75 for passing AOBEM Part I. This predicted probability of passing increased to .95 for examinees with Level 1 scores of 500. The predicted probability of failing is 1 minus the predicted probability of passing. Thus, the examinees with Level 1 scores of 400 are at 5 times greater risk of failing AOBEM Part I than those with scores of 500 (.25/.05).

which candidates took AOBEM Part I. By contrast, candidates typically took COMLEX-USA Level 1 seven years before they took AOBEM Part I. With a longer interval, more potential factors could change examinees' abilities and performances. In addition, by the time that candidates take COMLEX-USA Level 3, many are already training in postdoctoral programs (eg, emergency medicine residency programs for this sample) and are

Discussion

Scores from COMLEX-USA Level 1, Level 2-CE, and Level 3 revealed statistically significant positive moderate correlations with scores from AOBEM Part I. Scores on COMLEX-USA Level 3 were most highly correlated with AOBEM Part I scores, followed by COMLEX-USA Level 2-CE and Level 1 scores in terms of strength of correlation. This evidence supported the discriminative validity of COMLEX-USA to some extent. During the time between the COMLEX-USA series and AOBEM Part I, the year in which candidates took COMLEX-USA Level 3 was closest to the time frame in

Table 3.
Comparison of COMLEX-USA Scores Between
AOBEM Part I Pass Group and Fail Group (N=451)

Examination	Group, mean (SD)		t_{450}	P Value
	Pass (n=397)	Fail (n=54)		
COMLEX-USA Level 1	484.1 (68.6)	421.6 (57.5)	6.40	<.001
COMLEX-USA Level 2-CE	494.8 (75.5)	419.4 (69.1)	6.95	<.001
COMLEX-USA Level 3	510.4 (114.6)	394.2 (89.3)	7.16	<.001

Abbreviations: AOBEM, American Osteopathic Board of Emergency Medicine; CE, Cognitive Evaluation; COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA; SD, standard deviation.

Table 4.
Logistic Regression Model for the Log-odds of Passing AOBEM Part I as Predicted by COMLEX-USA Performance

Examination	Estimate	Standard Error	Wald χ^2	Pr$>\chi^2$
Logistic Regression Step 1				
COMLEX-USA Level 1	0.009	0.003	7.36	<.001
COMLEX-USA Level 2-CE	0.010	0.003	12.35	<.001
Logistic Regression Step 2				
COMLEX-USA Level 1	0.007	0.003	4.78	.03
COMLEX-USA Level 2-CE	0.006	0.003	3.42	.06
COMLEX-USA Level 3	0.006	0.002	11.06	<.001

Abbreviations: AOBEM, American Osteopathic Board of Emergency Medicine; CE, Cognitive Evaluation; COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination-USA; SD, standard deviation.

expected to demonstrate knowledge of clinical concepts and principles necessary for solving medical problems as independently practicing osteopathic generalists. In the present study, examinees' knowledge, skill, and ability to manage clinical problems in the unsupervised practice setting was most comparable to those of AOBEM Part I candidates. By contrast, COMLEX-USA Level 1 focuses more on scientific understanding of health and disease, often referred to as clinically applied foundational biomedical sciences and osteopathic principles. Therefore, we were not surprised that COMLEX-USA Level 1 scores were least correlated and Level 3 scores were most correlated with AOBEM Part I scores.

According to the multiple regression models, COMLEX-USA Level 1 and Level 2-CE scores, which some residency program directors use as part of the selection criteria for residents, together explained 30% of variance in AOBEM Part I scores for the sample of this study. This result is equivalent to .55 as a joint correlation between COMLEX-USA Levels 1 and 2-CE and AOBEM Part I performances. Adding COMLEX-USA Level 3 scores explained 7% of extra variation of AOBEM Part I scores. Considering that "the COMLEX-USA examination series is designed to assess the osteo-

pathic medical knowledge and clinical skills considered essential for osteopathic generalist physicians to practice osteopathic medicine without supervision,"[9(p7)] along with the fact that emergency medicine is not generally classified as a primary care discipline, it is understandable that the explained variance was not higher. These results also suggest that COMLEX-USA Level 1 and Level 2-CE could serve as effective and important partial criteria in predicting whether candidates pass or fail AOBEM Part I.

Differences in COMLEX-USA Level 1, 2-CE, and 3 scores were statistically significant among the examinees who passed and the examinees who failed AOBEM Part I. However, when all 3 COMLEX-USA scores were included in the logistic model to predict the odds of passing AOBEM Part I, Level 2-CE scores were not statistically significant. Even though COMLEX-USA Level 2-CE and Level 3 have different emphases, Level 3 was more similar in terms of clinical knowledge and application of principles and management of patient presentations to Level 2-CE relative to Level 1. Because of the similarity (ie, strong correlation) between Level 2-CE scores and Level 3 scores, the model picked the stronger predictor (Level 3). This logistic model may also apply to the individual level: residents can use their Level 3 scores to assess their chance of passing AOBEM Part I with a 95% confidence level. When the predicted chance of passing for a resident is low, the resident may take extra effort in preparing for AOBEM Part I.

Limitations

We identified a few limitations in this research, the first of which is straightforward: The sample in this study is restricted to the performances of emergency medicine residents. Whereas another study[6] showed a significant relationship between outgoing in-service examination performance and first-time success on AOBEM Part I, one may question how generalizable these results are to other specialty board examinations. We see no content-specific reasons to suggest that similar results would not be present between COMLEX-USA and other specialty board examination performances.

The second limitation is less obvious: The probabilities associated with passing AOBEM Part I were conditioned on the fact that candidates passed the COMLEX-USA series and were accepted into an emergency medicine residency program. That is, one would obtain a more accurate picture if residency program accepted and unaccepted statuses were available. With such data, we suspect that our statistical models would have more discriminatory power to predict AOBEM Part I performances from COMLEX-USA performances.

The third limitation is that this study did not control any other factors that might affect AOBEM Part I performance. For example, resident program–relevant variables such as the rank of the program, the size of the program, and the performance in the residency compared with the performance of peers (such as the ratings received by residents from program directors and performances on the osteopathic emergency medicine resident in-service examination) were not controlled.

Future Research

One direction for future research involves relating COMLEX-USA performances to additional measures of competence, including clinical patient outcome measures. In the context of this study, one can define a candidate as "competent" if the candidate passes all 3 parts of the AOBEM examination series. Second, we encourage research that relates COMLEX-USA performances to measures of competence in other medical specialties. Last, we believe that resident program–specific information (eg, rank or size of the program), the performance in the residency program, and candidates' demographic characteristics would improve the predictive power of future studies on this topic.

Conclusion

The present study empirically supports the predictive and discriminant validities of the COMLEX-USA series in relation to the AOBEM Part I certification examination. Residency programs may use COMLEX-USA Level 1 and Level 2-CE scores as part of the criteria used
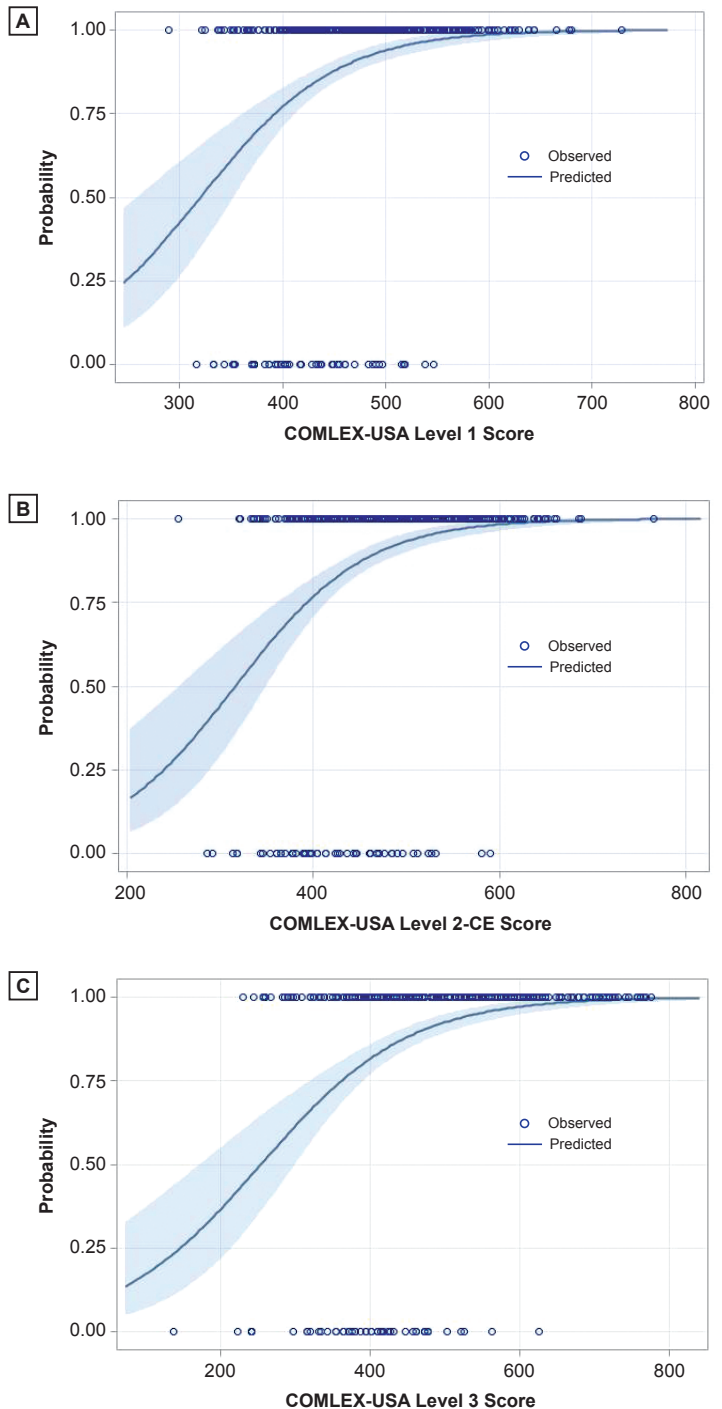
Figure.
Predicted probabilities for performance, with 95% confidence limits, on the American Osteopathic Board of Emergency Medicine Part I certification examination by scores on the (A) Comprehensive Osteopathic Medical Licensing Examination-USA (COMLEX-USA) Level 1, (B) COMLEX-USA Level 2-Cognitive Evaluation (CE), and (C) COMLEX-USA Level 3.

in selecting residents. Level 3 scores, though typically not available at the time of application, are actually a stronger predictor of performance on AOBEM Part I. Future researchers should choose measures that get as close as possible to measuring professional competence. Examples might include board certification, performance in practice assessments, and other clinical patient outcome measures.

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; 1999.

2. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945-949. doi:10.1016/S0140-6736(00)04221-5.

3. Kane MT. The validity of licensure examinations. *Am Psychol*. 1982;37(8):911-918.

4. 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.

5. Swanson DB, Sawhill A, Holtzman KZ, et al. Relationship between performance of the American Board of Orthopaedic Surgery Certifying Examination and scores on USMLE Steps 1 and 2. *Acad Med*. 2009;84(10 suppl):S21-S24.

6. Levy D, Dvokin R, Schwartz A, Zimmerman S, Li F. Correlation of the Emergency Medicine Resident In-Service Examination with the American Osteopathic Board of Emergency Medicine Part I. *West J Emerg Med*. 2014;15(1):45-50.

7. *Results of the 2012 NRMP Program Director Survey*. Washington, DC: National Resident Matching Program; August 2012. http://www.siumed.edu/oec/Year4/References /NRMP%20PDSurvey%202012.pdf. Accessed March 11, 2014.

8. Langenau EE, Pugliano G, Roberts WL. Competency-based classification of COMLEX-USA Cognitive Examination test items. *J Am Osteopath Assoc*. 2011;111(6):396-402.

9. National Board of Osteopathic Medical Examiners. *COMLEX-USA Bulletin of Information 2013-2014*. Conshohocken, PA: National Board of Osteopathic Medical Examiners; July 1, 2013. http://www.nbome.org/docs/comlexBOI.pdf. Accessed September 22, 2013.

Electronic Table of Contents

More than 110,000 individuals receive electronic tables of contents (eTOCs) for newly posted content to *The Journal of the American Osteopathic Association* website. To sign up for eTOCs and other *JAOA* announcements, visit http://www.jaoa.org/subscriptions/etoc.xhtml.