Medical Education

Original Article

Xiaolin Wang, PhD*, Hotaka Maeda, PhD, Brandon Craig, PhD, Tsung-Hsun Tsai, PhD, Jeanne M. Sandella, DO and Marie Fleury, DO, MBA

Meaningful use of COMSAE Phase 1 in preparation for COMLEX-USA Level 1

Abstract

Context: The Comprehensive Osteopathic Medical Licensing Examination of the United States (COMLEX-USA) is a three level national standardized licensure examination for the practice of osteopathic medicine. The Comprehensive Medical Self Assessment Examination (COMSAE) is a three phase self assessment tool designed to gauge the base knowledge and ability of candidates preparing for COMLEX-USA.

Objectives: To investigate how COMSAE Phase 1 (Phase 1) was used by candidates and how completing Phase 1 impacted their performance on the COMLEX-USA Level 1 (Level 1) examination.

Methods: Using data from the 2018–2019 administration of Level 1 and Phase 1 examinations, we counted the frequency of the unique Phase 1 forms taken by the candidates and calculated the correlation between the candidates' first attempt Phase 1 scores and the number of Phase 1 forms taken. We then calculated the correlation between the Level 1 scores and the Phase 1 scores. Next, we applied a multilevel regression model to examine the candidates' score improvement on the multiple Phase 1 forms taken. Finally, we investigated the effect of practicing through Phase 1 on the candidates' Level 1 performance using logistic regression models.

Results: The majority of candidates took one (2,414; 33.9%) to two (2,196; 30.8%) timed Phase 1 forms prior to the Level 1 examination. There was a significant negative correlation (r=−0.48, $t_{(6,505)}$=−44.05, p<0.001) between the candidates' first attempt Phase 1 scores and the number of Phase 1 forms taken. There was a strong and positive correlation (r=0.66 to 0.74, p<0.001) between Phase 1 and Level 1 scores. With other variables controlled, on average, candidates' Phase 1 scores increased 23.2 points on one attempt from the previous attempt. Having the most recent Phase 1 score controlled, a greater number of Phase 1 forms taken was associated with an improvement on the Level 1 performance.

Conclusions: The significant correlation between Phase 1 and Level 1 performance provided validity evidence for Phase 1. Moreover, our results suggested that candidates, especially those with lower performance on their initial Phase 1 attempt, might improve their Level 1 performance by taking multiple Phase 1 forms to monitor their academic improvement and gauge their readiness for Level 1.

Keywords: COMLEX-USA; COMSAE; medical education; preparation.

The Comprehensive Osteopathic Medical Licensing Examination of the United States (COMLEX-USA) is a three level national, standardized, high stakes licensure examination for the practice of osteopathic medicine. COMLEX-USA Level 1 (Level 1) assesses the knowledge of foundational biomedical sciences and other medical knowledge relevant to solving clinical problems and promoting and maintaining health in providing osteopathic medical care to patients [1]. It is administered in a timed setting and typically taken by the candidates of accredited osteopathic medical schools at the end of their second year. There are multiple test forms administered every test cycle. All forms are comparable in content and statistical specifications.

The Comprehensive Medical Self Assessment Examination (COMSAE) is a three phase self assessment designed to gauge the base knowledge and ability of candidates preparing for COMLEX-USA. Each COMSAE examination is presented in a format and structure similar to the

All authors are from the National Board of Osteopathic Medical Examiners.

*Corresponding author: Xiaolin Wang, PhD, National Board of Osteopathic Medical Examiners, Corporate Offices and National Center for Clinical Skills Testing, 8765 W. Higgins Rd, Suite 200, Chicago, IL 60631-4174, USA, E-mail: xwang@nbome.org
Hotaka Maeda, PhD, **Brandon Craig**, PhD, **Tsung-Hsun Tsai**, PhD, **Jeanne M. Sandella**, DO **and Marie Fleury**, DO, MBA, National Board of Osteopathic Medical Examiners, Corporate Offices and National Center for Clinical Skills Testing, Chicago, IL, USA

corresponding COMLEX-USA examination. COMSAE Phase 1 (Phase 1) is built from the same blueprint as Level 1 and is used as a self assessment tool to measure candidates' readiness to pass or achieve a certain score on Level 1 [2]. It has multiple test forms available for candidates or schools to purchase. Candidates or schools may choose to take the same test form multiple times, or they can order different forms. Schools can purchase any available forms, whereas candidates can purchase a limited number of forms.

Since COMSAE is designed as a self assessment tool for COMLEX-USA and the Phase 1 examination was ranked by 113 students among the most helpful preparation tools for Level 1 [3], we studied how Phase 1 was used by the candidates and how taking Phase 1 impacted their performance on the Level 1 examination. We hope the findings from this study can provide guidance to candidates regarding how their practice through Phase 1 can assist with their Level 1 preparation. In this study, we intended to answer the following questions: How many Phase 1 forms did candidates take prior to Level 1? What was the correlation between Level 1 and Phase 1 scores? Did Phase 1 performance improve as more forms were taken? Did the Phase 1 test taking experience help with the preparation for Level 1?

Methods

The study was reviewed and deemed exempt by the institutional review board of the National Board of Osteopathic Medical Examiners in July 2020.

Data

The Phase 1 data for this study was collected for the period from November 2017 to April 2019, a Phase 1 test window for students who planned to take the 2018–19 cycle of the Level 1 examination. We prepared the data by removing "outlier" scores from students who answered less than 90% of the questions and excluding tests completed under the untimed setting. Since candidates could repeat the same test form multiple times, we used their responses on the initial attempt for each unique form. In concert with the purpose of the study, only candidates who took Phase 1 prior to Level 1 were included.

All data analyses were conducted using R version 3.6.3 [4].

Number of Phase 1 forms taken

We counted the frequency of the unique Phase 1 forms taken by the candidates. We also calculated the correlation between the candidates' first attempt Phase 1 scores and the numbers of Phase 1 forms taken.

Correlation between Level 1 and Phase 1 scores

We then calculated the correlation between Level 1 scores and Phase 1 first attempt scores, last attempt scores, and mean scores across all forms.

Improvement of Phase 1 performance

Next, we examined the candidates' improvement on Phase 1. We modeled the relationship between the sequence of each unique form taken by a candidate (*sequence*) and the scores obtained on that form (*comsae.score*). This analysis was only based on the candidates who took more than two forms. Given that the scores were nested within the individual candidates, we calculated the intraclass correlation, which was 0.47. This value justified the application of a multilevel model. A random intercept model was specified as follows:

$$Comsae.score \sim sequence + year\_comsae + (1|ID)$$

As indicated, the candidates' Phase 1 scores were predicted by the exam level predictor *sequence* while controlling for the years spent in medical school when each Phase 1 form was taken (*year_comsae*). *Year_comsae* was computed based on the assumptions that all the candidates attended 4 year colleges and that schooling began on August 1 of the year they began attending the program. *Year_comsae* was computed by deducting the number of days left at school, which was calculated based on the candidates' school reported expected years of graduation, from the 4 year length of education and then divided by 365. $(1|ID)$ indicated a multilevel structure of the exam scores under individual candidates.

Improvement of the Level 1 performance

We investigated the effect of taking Phase 1 on candidates' Level 1 performance through a series of logistic regression models. The models described the relationship between the candidates' likelihood of passing Level 1 and a binary variable indicating whether additional Phase 1 forms were taken, given the same current Phase 1 performance. The models were:

- Model 1 (for *n.comsae*>=1): $logit(p.comlex) \sim comsae.score_1 + (n.comsae>1) + timegap$
- Model 2 (for *n.comsae*>=2): $logit(p.comlex) \sim comsae.score_2 + (n.comsae>2) + timegap$
- Model 3 (for *n.comsae*>=3): $logit(p.comlex) \sim comsae.score_3 + (n.comsae>3) + timegap$
- Model 4 (for *n.comsae*>=4): $logit(p.comlex) \sim comsae.score_4 + (n.comsae>4) + timegap$

In these models, *p.comlex* indicated the probability of passing Level 1, *comsae.score_i* indicated the *i*th-attempt Phase 1 score, $(n.comsae>I)$ was a binary variable indicating whether the candidate took more than *I* forms, where $I \in \{1, 2, 3, 4\}$, and *timegap* was the year difference between the first Phase 1 exam and the Level 1 exam. For example, Model 2 examined the difference in the Level 1 performance between the candidates who took exactly two Phase 1 forms (*n.comsae*=2) and those who took more than two forms (*n.comsae*>2), controlling for the second attempt Phase 1 score and the time gap between the first Phase

1 and the Level 1 examinations. Notably, each model in the series used a subset of the data from the previous model.

Results

Number of Phase 1 forms taken

Level 1 data were collected from 7,120 first time candidates who took the 2018–2019 administration of the examination between May 2018 and April 2019. Table 1 displays the frequency analysis results on the number of Phase 1 forms taken by candidates. The majority of candidates took one (2,414; 33.9%) to two (2,196; 30.8%) timed Phase 1 forms prior to the Level 1 examination. Additionally, a correlation analysis was conducted between the candidates' first attempt Phase 1 scores and the number of unique forms they took. A significant and negative Pearson's product-moment correlation (r = −0.48, $t_{(6,505)}=-44.05$, p<0.001) was found, indicating that candidates who obtained a higher first attempt score on Phase 1 tended to take fewer forms.
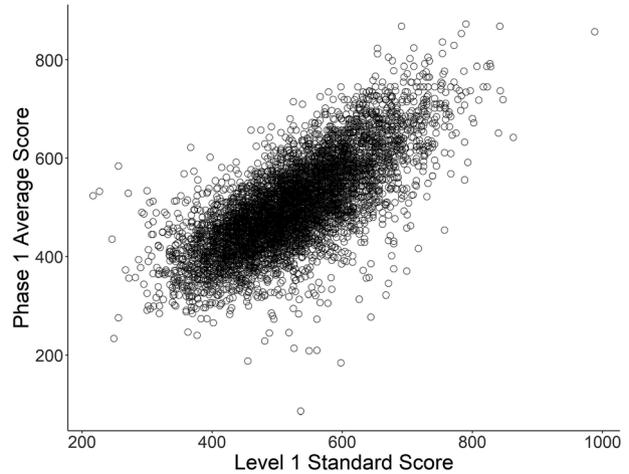
Correlation between Level 1 and Phase 1 scores

Table 2 reports the Pearson's product moment correlations between Level 1 scores and first attempt, last attempt, and means scores of all Phase 1 forms. All correlations were strong and significantly positive. The highest correlation occurred between the Level 1 scores and the mean Phase 1 scores (r=0.74, $t_{(6,505)}=87.51$, p<0.001), followed by last attempt Phase 1 scores (r=0.69, $t_{(6,505)}=76.86$, p<0.001).

Figure 1 shows the Level 1 scores by Phase 1 mean score. There is a clear positive relationship between the two sets of scores. A similar pattern was observed between the Level 1 scores and first/last attempt Phase 1 scores. Figure 2 shows the scatterplots of Level 1 scores and Phase 1 average scores for the candidates who took one, two, or three Phase 1 forms (i.e., the majority of candidates). The Pearson's product moment correlations were 0.69 ($t_{(2,412)}=47.37$, p<0.001) and 0.66 ($t_{(2,194)}=40.88$, p<0.001) for the candidates who took one and two Phase 1 forms. The correlation

Table 2: Pearson's product-moment correlation between Level 1 and Phase 1 scores.

	r	t	df	p-Value
Comsae.score_first	0.66	71.81	6,505	<0.001***
Comsae.score_last	0.69	76.86	6,505	<0.001***
Comsae.score_mean	0.74	87.51	6,505	<0.001***

***Indicates a significance level of 0.001.

Figure 1: Scatter plot of Level 1 standard scores and Phase 1 average scores (overall). The Pearson's product moment correlation between the Level 1 and Phase 1 average scores was 0.74 (p<0.001) based on all candidates who took at least one Phase 1 form.

was highest (i.e., r=0.81, $t_{(1,248)}=49.20$, p<0.001) for candidates who took three Phase 1 forms.

Improvement of Phase 1 performance

Figure 3 shows that first attempt scores were negatively correlated with the number of forms taken. While the group of candidates who only took one form (n=2,414) had a mean first attempt score of 546, the group of candidates who took five forms (n=145) had a mean score of 351. Within each subgroup, the mean Phase 1 score increased as another form was taken. We also observed that the mean score on the last forms taken was over 500 for all subgroups. In other words, the candidates seemed to target a score of 500 on Phase 1 as a threshold to stop taking

Table 1: Number of Phase 1 forms taken by candidates.

Number	0	1	2	3	4	5	6	Total
Frequency	613	2,414	2,196	1,250	501	145	1	7,120
Proportion	8.6%	33.9%	30.8%	17.6%	7.0%	2.0%	0.01%	100.0%
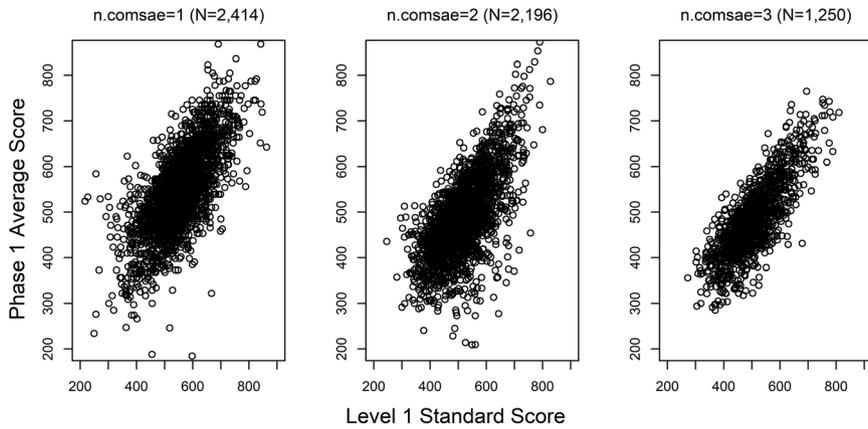
Figure 2: Scatter plot of Level 1 standard scores and Phase 1 average scores by number of Phase 1 forms taken. The Pearson's product-moment correlations were 0.69 (p<0.001), 0.66 (p<0.001) and 0.81 (p<0.001) for the three groups.

another form. The linear regression results of Phase 1 score improvement are displayed in Table 3. The sequence of the forms taken was a significant ($t_{(6,273)}$=29.35, p<0.001) predictor of Phase 1 scores. Specifically, with the timing of the test controlled, one unit increase in the sequence of the test was related to an increase of 23.2 points in Phase 1 score (i.e., the expected second attempt score would be 23.2 points higher than the first attempt score).
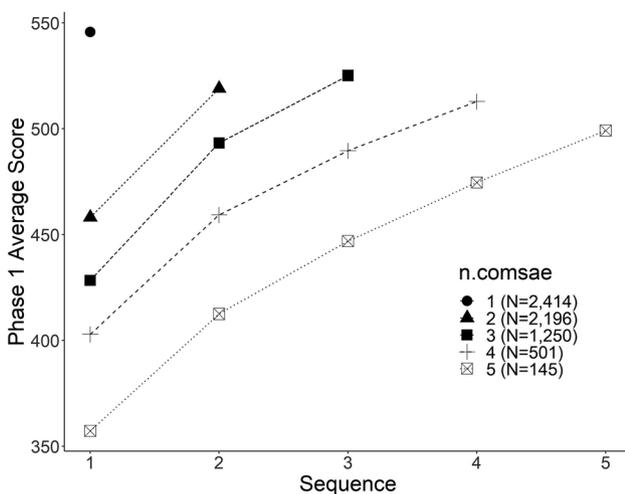
Improvement of the Level 1 performance

As shown in Table 4, taking extra Phase 1 forms showed either significantly positive effect (Models 1 and 3; p<0.05) or no significant effect (p=0.841 for Model 2 and p=0.053 for Model 4) on Level 1 performance. That means, for example, based on Model 1, holding other variables constant, the

Figure 3: Phase 1 mean score by number of forms taken and sequence. "n.comsae" indicates the number of Phase 1 forms taken prior to the Level 1 examination.

Table 3: Results of linear regression for Phase 1 practice effects on Phase 1 performance.

Fixed effects	Estimate	SE	t	df	p-Value
Intercept	144.11	8.00	18.01	5,205	<0.001***
Sequence[a]	23.17	0.79	29.35	6,273	<0.001***
year_comsae[b]	165.19	5.20	31.78	5,364	<0.001***

Random effects	Variance	SD
ID[c]	6,603	81.3
Residual	2,324	48.2

[a]"*sequence*" indicates the sequence of a Phase 1 form a candidate took; [b]"*year_comsae*" indicates the estimated year of time in medical school when a candidate took each Phase 1 examination; [c]"*ID*" indicates the individual candidates; ***indicates a significance level of 0.001. df, degree of freedom; SD, standard deviation; SE, standard error.

candidates who took more than one Phase 1 form (*n.comsae*>1) had an expected increase of 0.5 in the log odds of passing the Level 1 examination than those who took only one form (*n.comsae*=1). In comparison, based on Model 2, holding other variables constant, the candidates who took more than two Phase 1 forms (*n.comsae*>2) did not have significantly higher log odds of passing Level 1 than those who took exactly two forms (*n.comsae*=2).

Figure 4 exemplifies the probability curves for the two groups of candidates (*n.comsae*=1 vs. *n.comsae*>1) based on Model 1 when the time gap was fixed at its mean value of 0.21 (77 days). In this example plot, for the candidates who scored 400 on their first Phase 1 exam, those who stopped taking new forms had an 85% probability of passing the Level 1 examination, while those who took more forms had a 90% probability of passing Level 1. One can also notice in the plot (Figure 4) that candidates with lower scores benefited from taking extra forms more than candidates with higher scores.

Table 4: Results of linear regression for relationship between number of Phase 1 forms and Level 1 performance.

Model[c]	Inclusion criterion	Variable	Estimate	SE	z	p-Value
Model 1 (n=6,507)	n.comsae[a]>=1	Intercept	−5.98	0.43	−13.82	<0.001[c,***]
		n.comsae[a]>1 (yes)	0.53	0.14	3.76	<0.001[c,***]
		1st-attempt Phase 1 score	0.02	0.00	20.59	<0.001[c,***]
		timegap[b]	2.84	0.39	7.24	<0.001[c,***]
Model 2 (n=4,093)	n.comsae[a]>=2	Intercept	−6.93	0.53	−13.08	<0.001[c,***]
		n.comsae[a]>2 (yes)	0.03	0.13	0.20	0.841
		2nd-attempt Phase 1 score	0.02	0.00	17.79	<0.001[c,***]
		timegap[b]	5.01	0.66	7.61	<0.001[c,***]
Model 3 (n=1,897)	n.comsae[a]>=3	Intercept	−10.19	0.83	−12.30	<0.001[c,***]
		n.comsae[a]>3 (yes)	0.61	0.19	3.31	<0.001[c,***]
		3rd-attempt Phase 1 score	0.03	0.00	14.39	<0.001[c,***]
		Timegap[b]	3.23	1.08	2.98	<0.01[b]
Model 4 (n=647)	n.comsae[a]>=4	Intercept	−8.27	1.29	−6.43	<0.001[c,***]
		n.comsae[a]>4 (yes)	0.65	0.33	1.93	0.053
		4th-attempt Phase 1 score	0.02	0.00	7.82	<0.001[c,***]
		timegap[b]	0.01	1.43	0.01	0.996

[a]"n.comsae" is the number of unique Phase 1 forms taken prior to Level 1. [b]"timegap" is the gap in years between the first Phase 1 exam and the Level 1 exam. [c]Each model in the series used a subset of the data from the previous model, thus the sample size was the greatest for Model 1 (for candidates who took at least one Phase 1 form; n.comsae>=1) and the smallest for Model 4 (for candidates who took at least four Phase 1 forms; n.comsae>=4). SE, standard error. [***]Indicates a significance level of 0.001.

Discussion

There is scant research exploring how medical students use COMSAE self assessment tools to prepare for the COMLEX-USA examinations. The purpose of this study was to investigate the candidate experience with Phase 1 to determine whether and how it helped them prepare for the Level 1 examination.

The results showed a strong relationship between Phase 1 and Level 1 scores. Phase 1 and Level 1 are built from the same blueprint. Phase 1 questions are repurposed questions from the Level 1 examination and are presented with a similar format and structure. It is unsurprising to see a significant correlation between the two sets of scores. Jackson et al. [5] analyzed Phase 1 and Level 1 score information from 102 survey participants from the class of 2019 at the University of New England College of Osteopathic Medicine and found a similar correlation of 0.61 between participants' Level 1 scores and their scores on a Phase 1 form taken nearest to their Level 1 examination. However, they also found a low correlation of 0.25 between participants' Level 1 scores and their scores on a Phase 1 form taken earlier [5]. They then concluded that Phase 1 was not a meaningful preparation tool for Level 1 [5]. We argue that the low correlation was likely a measurement error caused by their small sample size. Contrary to Jackson et al.'s conclusion, our findings supported the predictive validity of Phase 1.
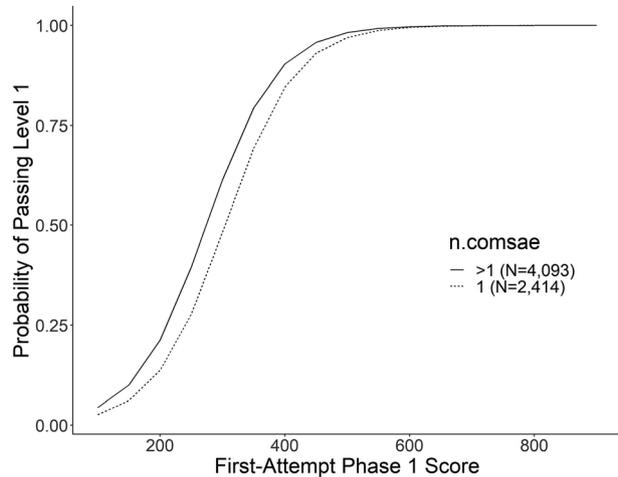
Figure 4: Probability of passing Level 1 for candidates who took one Phase 1 form vs. who took more than one form.

Most candidates in our study took one or two Phase 1 forms and the mean Phase 1 last attempt score was over 500 among all the subgroups taking different numbers of Phase 1 forms. We infer that students tended to continue taking the Phase 1 examination until they reached a score of 500 or above; thus, it seems that the candidates used Phase 1 as a tool to monitor their academic progress until they felt confident that they would pass or reach a certain score (i.e., a target score for the application of certain competitive residency or training programs) on Level 1.

For the candidates in our study with the same recent Phase 1 performance, taking more forms was generally related to an increase in the probability of passing Level 1. However, note that this relationship does not indicate causality. It may not be that taking Phase 1 alone helped with better Level 1 performance. Instead, the candidates chose to keep studying, monitored their progress through more Phase 1 forms, and achieved better Level 1 results, as suggested by the improvement in Phase 1 scores. We encourage future candidates, especially those with lower performance on their initial Phase 1 attempts, to utilize Phase 1 to gauge their readiness for Level 1 and monitor their improvement.

A limitation to our study is that we treated all Phase 1 test taking in the same way, with the belief that candidates might benefit from exposure to Phase 1 content and format, regardless of why they took the examination. We acknowledge that some schools use Phase 1 as a barrier to qualify students to take Level 1. The reasons why candidates took Phase 1 can be another parameter we explore in a future study. Another practical consideration is that Phase 1 is a variant examination, allowing different modes of timing and proctoring and involving different levels of candidate motivation. Thus, relying on Phase 1 scores alone to identify candidates with "outlier" scores for early intervention may not be feasible. Future studies can collect more data, such as demographic information (e.g., gender, ethnicity, or socioeconomic status) and classroom performance data, so that more accurate predictions of Level 1 performance can be achieved. Further, we applied the data from Phase 1 and Level 1 examinations built from the original test blueprint [6], which was composed of nine patient presentation domains and six physician tasks. From the 2019–2020 administration, the Level 1 examination adopted a new test blueprint [7], which consists of seven competency domains and 10 clinical presentations. To examine the impact of the change in test blueprint on the results, we plan to reexamine the findings in May 2021, when we have sufficient data.

Conclusions

This study examined how candidates used COMSAE Phase 1 as a self assessment tool to prepare for the COMLEX-USA Level 1 examination. The significant correlation between Phase 1 and Level 1 performance provides validity evidence for Phase 1. Moreover, results suggested that candidates, especially those with lower performance on their initial Phase 1 attempt, might improve their Level 1 performance by taking multiple Phase 1 forms.

References

1. COMLEX-USA Level 1. https://www.nbome.org/exams-assessments/comlex-usa/comlex-usa-level-1/ [Accessed January 4 2021].
2. COMSAE Phase 1. https://www.nbome.org/exams-assessments/comsae/traditional/new-comsae-phase-1/ [Accessed January 4 2021].
3. Vora A, Maltezos N, Alfonzo L, Hernandez N, Calix E, Fernandez MI. Predictors of scoring at least 600 on COMLEX-USA Level 1: successful preparation strategies. J Am Osteopath Assoc 2013;113: 164–73.
4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer software]; 2020. Available from: https://www.R-project.org/.
5. Jackson F, Duane E, Harmon R, Kollar RA, Rainville NM, Smith RM. Resources that improve medical board licensing examination performance. Cureus 2019;11:e5927.
6. Blueprint. COMLEX-USA Level 1. https://www.nbome.org/exams-assessments/comlex-usa/comlex-usa-level-1/blueprint/old-blueprint-comlex-usa-level-1/ [Accessed January 4 2021].
7. Blueprint. COMLEX-USA Level 1. https://www.nbome.org/exams-assessments/comlex-usa/comlex-usa-level-1/blueprint/ [Accessed January 4 2021].