Background: The National Board of Osteopathic Medical Examiners uses a standardized timing format for all osteopathic medical candidates across all stations in the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE). Examination standardization, and therefore validity, is weakened if time administered affects student performance.

Objective: To investigate the relationship between the amount of time students spent with standardized patients (SPs) and their clinical skill component scores.

Methods: Data were analyzed from 249 candidates in 21 test sessions from the 2007-2008 testing cycle. Trained SPs recorded the time each examinee entered and exited each encounter. Outcome measures of the examination included pass (1 if pass; 0 if fail) performance outcome, pass by domain (biomedical/biomechanical and humanistic), and pass by component scores (data gathering, osteopathic manipulative treatment, and patient note scores). Analysis of variance (ANOVA), followed by pairwise comparisons, was used to compare encounter times across cases.

Results: A total of 2988 SP encounters were analyzed. Examinees spent an average 12.1 minutes in each patient encounter. Candidates spent less time with cases classified as health promotion and disease prevention than with cases that primarily involved a problem visit, but amount of time spent with cases representing various content areas was not significantly different ($P>.05$). On average, candidates spent the least amount of time (11.95 minutes) with gastrointestinal cases and the most time (12.26 minutes) with patients who had complaints relating to the neuromusculoskeletal system. A trivial relationship was shown between the continuous variable encounter time as the average amount of time a student spends in the encounter and the dichotomous clinical skill outcome of the examination ($r=.03$). The relationship between these 2 variables was consistent when examined separately for the humanistic domain ($r=-.10$) and the biomedical/biomechanical domain ($r=.12$). The strongest correlation was between encounter time and the data-gathering component score ($r=.28$).

Conclusion: Time spent in the encounter was not influential in overall candidate performance in the COMLEX-USA Level 2-PE. Candidates spent a similar amount of time with cases of different content areas. This finding suggests that 14 minutes is sufficient for this osteopathic high-stakes performance evaluation.

The National Board of Osteopathic Medical Examiners administers the Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation (COMLEX-USA Level 2-PE). This performance-based evaluation of clinical skills is to be taken by osteopathic medical students in the United States. Candidates are primarily third-year (late in the academic year) and fourth-year osteopathic medical students. Level 2-PE assesses clinical skills in 2 domains: (1) the humanistic domain, which evaluates patient-physician communication, interpersonal skills, and professionalism and (2) the biomedical/biomechanical domain, which includes medical history completion and physical examination skills, osteopathic principles and practice (including osteopathic manipulative treatment [OMT]), and written communication skills. Performance standards must be met in both domains for a candidate to pass the examination.

Each candidate is given 14 minutes to evaluate and treat a standardized patient (SP) and an additional 9 minutes to complete a postencounter SOAP note (subjective, objective, assessment, plan). Pilot testing and a previous study have shown that 14 minutes is adequate to complete the tasks required for the case when the examinee is asked to “Evaluate and treat the patient as you see fit.”
Previous investigations\textsuperscript{5,6} reported weak positive relationships between time spent in the encounter and both interpersonal skills and data-gathering scores on high-stakes clinical skills examinations, suggesting that spending more time with the SP might increase performance scores in these areas. However, these studies\textsuperscript{5,6} did not examine student performance on the written communication task as a function of time and were not specific to the osteopathic assessment experience.

To administer a standardized format to all candidates across all stations, COMLEX-USA Level 2-PE strictly adheres to the allotted 14 minutes per station. Standardization, and therefore validity, is weakened if the time administered to examinees affects their performance. The intent of the present study was to investigate the relationship between the amount of time a candidate spends with an SP and his or her performance within the clinical skill component.

The blueprint for COMLEX-USA Level 2-PE was based on typical medical complaints of patients seen by primary care physicians, varying in age, gender, race, ethnicity, and nature of the chief complaint as reported by the National Ambulatory Medical Care Survey (NAMCS).\textsuperscript{7} The blueprint was also developed to emphasize the uniqueness of the osteopathic experience,\textsuperscript{8} and therefore, to be appropriate for osteopathic medical students preparing for entry into graduate medical education. To fit the test blueprint, cases in COMLEX Level 2-PE vary by age, sex, race, and nature of the chief complaint. The validity of the examination format is partially dependent on candidates’ ability to complete the tasks within the time constraints imposed.

The purpose of the present study is to investigate the amount of time spent by candidates in the SP encounter. We also sought to identify how encounter time affects COMLEX-USA Level 2-PE performance. Encounter time based on case content will also be investigated.

\section*{Methods}

A random selection of test sessions was drawn from the 2007-2008 testing cycle. Entire sessions were reviewed to represent complete test forms and also to ensure that all content categories were represented. When candidates take the COMLEX-USA Level 2-PE, they give permission for their deidentified data to be used for research purposes. Therefore, institutional review board approval was not sought for the current study. Candidates were deidentified for analysis, and results of the present study are reported in aggregate form to maintain confidentiality.

Data were collected on encounter timing and scores from 21 test sessions in which 249 candidates rotated through 12 SP cases resulting in 2988 patient-physician encounters. Data were collected across different case categories: cardiac, respiratory, gastrointestinal, neuromusculoskeletal, and other complaints. Data were also collected for the classifications of chronic issues, acute issues, and health promotion and disease prevention visits.

Test sessions represented candidates from 22 osteopathic medical schools. The majority were white (75\%), women (57\%), and spoke English as their primary language (90\%). The majority of examinees were first-time test takers (93\%).

\subsection*{Level 2-PE Testing and Scoring}

When the encounter begins for COMLEX-USA Level 2-PE, the candidate is given the prompt “You may begin your clinical encounter,” at which time the student has 14 minutes to complete the encounter. The candidate is not required to enter the room immediately—some choose to spend time reviewing the medical record before entering. Candidates must leave the examination room at the end of the 14-minute SP encounter. They may leave the room at their discretion before the 14-minute allotment; however, once the candidate leaves the encounter, reentry into the examination for any reason is prohibited, regardless if the 14 minutes allotted have not yet expired. Candidates must complete 12 patient encounters during the examination.

For the present study, encounter time was defined as the difference between the recorded time a candidate entered and exited the room. The SPs viewed videos of the encounters and noted the time that the encounter began by recording the time the “You may begin the clinical encounter” announcement played, and noted the encounter end time as the time that the candidate crossed the threshold to exit the room. The difference in time is called the “encounter time.” The recorders begin recording the encounters 30 to 60 seconds before the encounters, which can account for an exit time in excess of 14 minutes. All SPs had been trained previously on the use of the encounter viewing module and were standardized as to what to record. Recordings were made on paper and then 2 SPs individually and separately transcribed the recordings into a spreadsheet to ensure reliability of the transfer of that data.

The SPs score candidates on the humanistic domain using the Global Patient Assessment (GPA) tool.\textsuperscript{1,4} The GPA is scored on a Likert-type rating scale to record candidate behaviors related to patient-physician communication, interpersonal skills, and professionalism. Encounter-level ratings are the individual ratings given for each encounter and each component. The GPA scores from all encounters are averaged to determine the candidate’s final humanistic domain score.

The biomedical/biomechanical domain is scored on a percentage scale computed from the candidate’s weighted data-gathering, postencounter written note, and OMT scores, as follows:

\begin{itemize}
\item \textbf{Data gathering}—The data-gathering score is a percentage based on case-specific checklist items reflecting medical history asked and physical examination maneuvers performed,
\end{itemize}
which are recorded by trained SPs. The candidate’s data-gathering mean is computed from his or her encounter-level scores.

Postencounter note—The postencounter note, in SOAP note format, is rated by a trained osteopathic medical examiner using a Likert-type scale and scoring rubric. The candidate’s mean SOAP note rating is computed from his or her encounter-level scores.

OMT—Using a rubric, trained osteopathic physicians rate candidates on their use of osteopathic principles and practice, including OMT, in 25% to 40% of the encounters. The candidate’s mean OMT rating is computed from his or her encounter-level scores.

The biomedical/biomechanical domain score is computed within a compensatory scoring framework, where poor performance in 1 of the 3 clinical skill areas can be compensated for by better performance in 1 or more of the other clinical skill areas.

Statistical Analysis
Analysis of variance (ANOVA), followed by pairwise comparisons, was used to compare encounter times across cases. Statistical significance was set at \( P < .05 \) to control for type I errors for omnibus \( F \) and pairwise comparisons. Descriptive statistics are also provided.

Results
Time spent in the encounter is presented in Table 1. The mean (standard deviation [SD]) time spent in the patient encounter was 12.1 (1.0) minutes (range, 8.5-13.7 minutes). No candidate in the present study spent the entire 14 minutes in the examination room.

Descriptive statistics for time in the encounter by case presentation are provided in Table 2. The least amount of time among candidates was spent on gastrointestinal cases (mean [SD], 11.95 [1.16] minutes). In comparison, the maximum amount of time spent was on neuromusculoskeletal cases (mean [SD], 12.26 [1.18] minutes). The main effect for mean encounter time spent under case-specific SP presentations was statistically significant (\( F_{4,1299} = 3.07, P = .015 \)).

Follow-up pairwise comparisons between means are shown in Table 3 with effect sizes. The mean difference is the first case presentation mean minus the second case presentation mean from Table 2, where the sign of the mean difference reflects the larger mean. No pairwise comparison was statistically significant, indicating there is some weighted combination of case presentation means to account for the statistically significant main effect. Effect sizes accompanying each pairwise comparison shows approximately one-quarter SD separates the means for cardiovascular-neuromusculoskeletal (-0.20), respiratory-neuromusculoskeletal (-0.22), gastrointestinal-neuromusculoskeletal (-0.26), and gastrointestinal-other (-0.21) cases. No other pairwise comparison showed practical importance.

Descriptive statistics for time in the encounter by case presentation axis are presented in Table 4. Candidates, on average, spent less time with cases developed as health promotion and disease prevention (mean [SD], 11.58 [1.41] minutes) than with cases that primarily involved a problem visit. A problem visit case is classified as either acute (mean [SD] encounter time 12.18 [1.05] minutes) or chronic (mean [SD] encounter time, 12.34 [1.09] minutes). The test of the main effect for case axes was statistically significant (\( F_{2,749} = 28.06, P < .001 \)).
Pairwise comparisons between means are shown in Table 5 with accompanying effect sizes and are interpreted in the same way as stated above. Results show a statistically significant larger mean encounter time for acute cases compared to health promotion and disease prevention cases (mean difference, 0.60). The effect size (0.48) for this pair shows approximately one-half SD separates these 2 means. Results show a statistically significant larger mean encounter time for chronic cases compared to health promotion and disease prevention cases (mean difference, 0.76). The effect size (0.61) for this pair shows more than one-half SD separates these 2 means. Differences in acute and chronic case means were not statistically significant and showed little practical importance in their effect size.

The first 3 correlations shown in Table 6 reflect the extent of the relationship between encounter time and the clinical skill outcome of examinees. The correlation between encounter-time and the overall clinical skill outcome of examinees is trivial ($r = 0.03$), indicating little relationship between the amount of time an examinee spends in the encounter and his or her clinical skill outcome. A similar result was shown for the relationship between encounter time and clinical skill outcome of the humanistic domain among examinees ($r = -0.10$), and the relationship between encounter time and biomedical/biomechanical domain clinical skill outcome of examinees ($r = 0.12$).

A significant correlation was shown between time spent in the encounter and data gathering component scores ($r = 0.28$, $P < 0.01$), indicating that time spent in the encounter explains approximately 7.8% of the variance in data-gathering scores (Table 6). No other clinical skill component score resulted in a correlation with time spent in the encounter of statistical or practical importance.

**Comment**

According to Langenau et al, 19.5% of candidates in the 2008-2009 test cycle either disagreed or strongly disagreed with the statement, “The time allowed to complete the encounter with the SP was sufficient” and 18.7% responded similarly regarding time to complete the written SOAP note. Candidates seem concerned about time management and perceive there is not enough time provided to spend with the SP or to write the postencounter note. Findings from this study are reflective of national examination administration standards, therefore generalizations may not completely address time management concerns because examinees are limited to the 14-minute standardized time protocol in this examination administration. However, findings from this study provide supportive evidence for a 14-minute limit for the patient-physician encounter in a standardized administration of a national licensure examination given that within this restriction, overall candidate performance is not influenced by encounter time.

**Performance and Timing**

The present study demonstrates that time is not an influential factor with respect to performance in the overall examination, the humanistic domain, the biomedical/biomechanical domain, or 2 of the 3 biomedical/biomechanical domain subcomponent scores (ie, SOAP note and OMT). Disaggregating the biomedical/biomechanical domain into its skill areas reveals a positive correlation between time and performance only for the data-gathering skill area. This associ-
This strategy may confound results of the present investigation. Further investigation into how time is used in an osteopathic clinical skills assessment is required to explore this relationship between encounter time and data-gathering scores.

Candidates who leave the encounter before the 14-minute time limit has expired increase the 9 minutes already allotted for the written note task, which could be perceived as an advantage to this task. Considering that all candidates are given the same opportunity to manage their time, choosing to take time away from the SP encounter does not appear to improve performance because there is no statistically significant correlation between encounter time and the SOAP note overall score.

It could be argued the candidate spends more time with patients to perform more in-depth OMT techniques. However, results show a lack of a relationship between encounter time and OMT scores.

Candidates are evaluated on patient-physician communication, interpersonal skills, and professionalism within the humanistic domain. Chambers et al. showed a positive relationship between longer encounters and higher communication skills scores that was not shown in the present study. In practice, patients who felt that they spent more time than expected with their physicians were more satisfied with their visits. That positive correlation between longer encounter times and scores on the humanistic domain was not demonstrated within this cohort of candidates taking the COMLEX-USA Level 2-PE. There may be qualities defined in the GPA tool that are time independent. A study looking specifically at these targeted skills would be of interest.

Under ideal conditions, a candidate would simply leave the examination room once sufficient information is gathered from the patient. However, there are several variables that could potentially affect encounter time that were not investigated in the present study. For example, perhaps the encounter is rushed because the candidate leaves early because of an illness or because the candidate is hoping to achieve an advantage by spending more time on SOAP note documentation. Conversely, a candidate may elect to spend additional time with the SP because the candidate realizes that there is remaining time to ask additional questions or perform additional physical examination maneuvers that may or may not be related to the patient’s complaint. Another possibility is that because a candidate is not permitted to leave and reenter the examination room, the candidate artificially prolongs encounter time by remaining in the room after completing the history and physical examination tasks, ensuring opportunity to ask additional questions if necessary. Each of these possible scenarios may confound the findings of the present study. Given these possibilities, one could imagine how encounter time in and of itself does not reveal how that time was effectively spent.

| Table 5. Candidate Performance on COMLEX-USA Level 2-PE: Case Presentation Axis Pairwise Comparison Mean Differences and Effect Sizes* (N=2988) |
|---------------------------------|-----------------|--------------|
| Presentation Axis Pairwise Comparison | Mean Difference | Effect Size |
| Acute-chronic                   | -0.16           | -0.15        |
| Acute-disease prevention        | 0.60*           | 0.48         |
| Chronic-disease prevention      | 0.76*           | 0.61         |

*Pairwise comparisons were computed using Tukey’s method. Effect sizes were computed using Hedge’s g.

Abbreviation: COMLEX-USA Level 2-PE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation.

| Table 6. Candidate Performance on COMLEX-USA Level 2-PE: Correlations Between Time Spent With Patients and Performance Scores (N=2988) |
|---------------------------------|-----------------|--------------|
| Clinical Skill Outcome*         | Correlation     |
| Both domains                    | 0.03            |
| Humanistic domain               | -0.10           |
| Biomedical/biomechanical domain | 0.12            |
| Clinical component scores†      |                 |
| SOAP note                       | 0.09            |
| Osteopathic manipulative assessment | -0.05          |
| Data gathering†                  | 0.28†           |
| Global Patient Assessment tool† | 0.09            |

†The clinical skill examination results in a binary outcome (1 if pass, 0 if fail). Encounter time is a continuous measure. Correlations between each of the three clinical skill outcomes and encounter time are point biserial correlation coefficients. Correlations between clinical skill component scores and encounter time are Pearson correlation coefficients. SOAP notes and osteopathic manipulative treatment are rated by physician raters and scored on a percentage metric. Data gathering is recorded by standardized patients and scored on a percentage metric. Data gathering is recorded by standardized patients and scored on a scale ranging from 1 to 9.

Abbreviations: COMLEX-USA Level 2-PE, Comprehensive Osteopathic Medical Licensing Examination-USA Level 2-Performance Evaluation; SOAP, subjective, objective, assessment, plan.
Case Characteristics and Timing

On average, candidates use 12 of the 14 minutes provided for each patient encounter. Similar to findings of previous studies, candidates did not spend more time with cases of differing characteristics. Although results yielded a case presentation main effect, follow-up pairwise comparisons between the 3 case presentations did not show differences between simple means; therefore, some weighted combination of case presentations contributed to the main effect. For example, candidates spend a similar amount of time with a patient with a respiratory complaint as with a patient with a musculoskeletal complaint.

Interestingly, candidates spend less time with SPs in cases that focus on health promotion and disease prevention when compared to acute or chronic problem visits. These cases usually rely more on the candidate’s ability to provide counseling to the patient about primary preventive measures in healthcare. Candidates may spend less time with these cases because they perceive an insufficient complaint to remedy, or perhaps feel they lack the skills necessary to discuss preventive care. Some recent studies have investigated candidates’ perception of the relevance of prevention counseling and found that it was positively predicted by factors such as interest in primary care, school attended, interest in prevention, female gender, nonwhite race, and having healthy personal practices.

It is not clear whether these factors would result in shorter encounter times, but further investigation could help elucidate these relationships.

Limitations

A random sample of 2988 encounters was reviewed, representing only 6.1% of the encounters for the entire 2007-2008 testing cycle. Also, generalization is based on content presentation, not case specific. Cases were not identified by age, sex, or the SP who participated in the encounter—all of which could possibly affect encounter time. Another limitation is first-time and repeat test takers were not separated in the analysis. It is conceivable that these cohorts may use their time differently. Also, we did not attempt to evaluate time spent on discrete tasks regarding history taking, physical examination, and OMT. Finally, although individual schools may administer their own clinical skills examinations to reflect the nature and timing of COMLEX-USA Level 2-PE, findings of the present study should not be generalized to school-based clinical skills examinations. Future investigations could involve analyzing time spent in simulation encounters at osteopathic medical schools based on the purpose of the assessment (eg, formative vs summative assessments).

Conclusion

The present study supports that 14 minutes is sufficient for the patient-physician encounter. The candidate’s clinical skills assessment on COMLEX-USA Level 2-PE is not likely to be affected by the amount of time spent with the SP.

Acknowledgments

We thank Kristie Lang at the National Board of Osteopathic Medical Examiners for her assistance with collecting and organizing the timing data and John R. “Jack” Boulet, PhD, of the Foundation for Advancement of International Medical Education and Research for his critical review of the manuscript.

References


Sandella et al • Medical Education