

Philadelphia College of Osteopathic Medicine

DigitalCommons@PCOM

---

PCOM Scholarly Papers

---

5-13-2021

## Examining concurrent validity between COMLEX-USA Level 2-Cognitive Evaluation and COMLEX-USA Level 2-Performance Evaluation.

Brandon Craig

Xiaolin Wang

Jeanne Sandella

Tsung-Hsun Tsai

David Kuo

Philadelphia College of Osteopathic Medicine, davidku@pcom.edu

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.pcom.edu/scholarly\\_papers](https://digitalcommons.pcom.edu/scholarly_papers)



Part of the [Medicine and Health Sciences Commons](#)

---

### Recommended Citation

Craig, Brandon; Wang, Xiaolin; Sandella, Jeanne; Tsai, Tsung-Hsun; Kuo, David; and Finch, Charles, "Examining concurrent validity between COMLEX-USA Level 2-Cognitive Evaluation and COMLEX-USA Level 2-Performance Evaluation." (2021). *PCOM Scholarly Papers*. 2111.

[https://digitalcommons.pcom.edu/scholarly\\_papers/2111](https://digitalcommons.pcom.edu/scholarly_papers/2111)

This Article is brought to you for free and open access by DigitalCommons@PCOM. It has been accepted for inclusion in PCOM Scholarly Papers by an authorized administrator of DigitalCommons@PCOM. For more information, please contact [library@pcom.edu](mailto:library@pcom.edu).

---

**Authors**

Brandon Craig, Xiaolin Wang, Jeanne Sandella, Tsung-Hsun Tsai, David Kuo, and Charles Finch

Brandon Craig\*, PhD, Xiaolin Wang, PhD, Jeanne Sandella, DO, Tsung-Hsun Tsai, PhD, David Kuo, DO and Charles Finch, DO

# Examining concurrent validity between COMLEX-USA Level 2-Cognitive Evaluation and COMLEX-USA Level 2-Performance Evaluation

<https://doi.org/10.1515/jom-2021-0007>

Received January 21, 2021; accepted April 23, 2021;

published online May 13, 2021

## Abstract

**Context:** The Comprehensive Osteopathic Medical Licensing Examination of the United States of America (COMLEX-USA) is a three level examination used as a pathway to licensure for students in osteopathic medical education programs. COMLEX-USA Level 2 includes a written assessment of Fundamental Clinical Sciences for Osteopathic Medical Practice (Level 2-Cognitive Evaluation [L2-CE]) delivered in a computer based format and separate performance evaluation (Level 2-Performance Evaluation [L2-PE]) administered through live encounters with standardized patients. L2-PE was designed to augment L2-CE. It is expected that the two examinations measure related yet distinct constructs.

**Objectives:** To explore the concurrent validity of L2-CE with L2-PE.

**Methods:** First attempt test scores were obtained from the National Board of Osteopathic Medical Examiners database for 6,639 candidates who took L2-CE between June 2019 and May 2020 and matched to the students' L2-PE scores. The sample represented all colleges of osteopathic medicine and 97.5% of candidates who took L2-CE during the complete 2019–2020 test cycle. We calculated disattenuated correlations between the total score for L2-CE, the L2-CE scores for the seven competency domains

(CD1 through CD7), and the L2-PE scores for the Humanistic Domain (HM) and Biomedical/Biomechanical Domain (BM). All scores were on continuous scales.

**Results:** Pearson correlations ranged from 0.10 to 0.88 and were all statically significant ( $p < 0.01$ ). L2-CE total score was most strongly correlated with CD2 (0.88) and CD3 (0.85). Pearson correlations between the L2-CE competency domain subscores ranged from 0.17 to 0.70, and correlations which included either HM or BM ranged from 0.10 to 0.34 with the strongest of those correlations being between BM and L2-CE total score (0.34) as well as between HM and BM (0.28). The largest increase between corresponding Pearson and disattenuated correlations was for pairs of scores with lower reliabilities such as CD5 and CD6, which had a Pearson correlation of 0.17 and a disattenuated correlation of 0.68. The smallest increase in correlations was observed in pairs of scores with larger reliabilities such as L2-CE total score and HM, which had a Pearson correlation of 0.23 and a disattenuated correlation of 0.28. The reliability of L2-CE was 0.87, 0.81 for HM, and 0.73 for BM. The reliabilities for the L2-CE competency domain scores ranged from 0.22 to 0.74. The small to moderate correlations between the L2-CE total score and the two L2-PE support the expectation that these examinations measure related but distinct constructs. The correlations between L2-PE and L2-CE competency domain subscores reflect the distribution of items defined by the L2-PE blueprint, providing evidence that the examinations are performing as designed.

**Conclusions:** This study provides evidence supporting the validity of the blueprints for constructing COMLEX-USA Levels 2-CE and 2-PE examinations in concert with the purpose and nature of the examinations.

**Keywords:** cognitive evaluation; COMLEX-USA, Level 2, medical education; performance evaluation; test scores.

The Comprehensive Osteopathic Medical Licensing Examination of the United States of America (COMLEX-USA) is a series of standardized assessments used in part to fulfill licensure requirements for the practice of osteopathic

---

\*Corresponding author: Brandon Craig, 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: bcraig@nbome.org

Xiaolin Wang, PhD and Tsung-Hsun Tsai, PhD, National Board of Osteopathic Medical Examiners, Chicago, IL, USA

Jeanne Sandella, DO, National Board of Osteopathic Medical Examiners, Philadelphia, PA, USA

David Kuo, DO, The Philadelphia College of Osteopathic Medicine, Philadelphia, USA

Charles Finch, DO, Midwestern University, Glendale, AZ, USA

medicine [1]. COMLEX-USA is comprised of four separate examinations spanning three progressive levels. Level 2 of COMLEX-USA, which is typically taken by students during their third or fourth year of medical school, is separated into a Cognitive Evaluation (L2-CE) and a Performance Evaluation (L2-PE). The L2-CE is a computer based, multiple choice assessment with 352 items that measures the application of knowledge in clinical and foundational biomedical sciences and osteopathic principles integrated with related physician competencies [2]. A passing grade on L2-CE is based on a single score, although subscores aligned with the blueprint dimensions are reported. L2-PE is a patient presentation based assessment of fundamental clinical skills. It requires that candidates demonstrate competency when they are presented with 12 standardized patient encounters [3]. A pass/fail score is reported for two domains: the Humanistic Domain (HM), which measures physician-patient communication, interpersonal skills, and professionalism; and the Biomedical/Biomechanical Domain (BM), which measures history and physical examination, documentation skills, and the performance of osteopathic manipulative treatment. Passing of the L2-PE is compensatory within domains but not across domains. Candidates must pass both domains on the same administration to pass the L2-PE.

The L2-PE was first administered in 2004 with the goal of augmenting L2-CE and to assess additional competencies required to provide patient care in supervised graduate medical education settings [4]. The motivation for adding this additional assessment to the COMLEX-USA was acknowledgment of the limitations of what can be measured with traditional multiple choice assessments. While the L2-CE is well suited for measuring candidates' medical knowledge and clinical reasoning skills, it is less adept at measuring clinical skills such as interpersonal skills, communication, hands on physical examination, or osteopathic manipulative treatment. The L2-PE was designed to measure these clinical skills [5] and help fulfill the mission of the National Board of Osteopathic Medical Examiners (NBOME) "to protect the public by providing the means to assess competencies of osteopathic medicine and related health care professions" [1].

While published research on L2-PE has supported the validity of the examination for use in determining candidates' competency to provide supervised patient care [4–7], no published research has explored concurrent validity between L2-CE and L2-PE. Given the similarities between these examinations in terms of the time at which they are taken by medical students and commonalities in the master blueprint along with expected differences due to the types of examinations, a study exploring the relationship between L2-PE and L2-CE is essential to provide

evidence of validity supporting the requirement that osteopathic medical school students demonstrate their knowledge and application of fundamental clinical skills for osteopathic medical practice on both assessments.

All examinations in the COMLEX-USA series share the master blueprint based on the same two dimensions, labeled as competency domains and clinical presentations. The seven competency domains (CD1 through CD7) and 10 clinical presentations are identical for all four examinations; however, the percentage of items aligned with each varies by examination. Table 1 shows the minimum percentages of items required for each competency domain for the L2-CE, L2-PE HM, and the L2-PE BM. The goal of this study was to examine the relationships between L2-PE and L2-CE by correlating the scores on HM and BM domains with the L2-CE total score and the L2-CE subscores for CD1 through CD7.

## Methods

This study design was reviewed by the Institutional Review Board (IRB) of the NBOME and deemed exempt. All analyses were conducted in R version 3.6.3.

Data from 6,639 candidates who took L2-CE between June 2019 and May 2020 were obtained from the NBOME database and matched to their L2-PE scores. Only first attempt test scores were included in this analysis. This sample represented all colleges of osteopathic

**Table 1:** Minimum percentage of items on the COMLEX-USA L2-CE, L2-PE Humanistic (HM) Domain, and L2-PE Biomedical/Biomechanical (BM) Domain exams.

| Competency domains  | Minimum percentage |       |       |
|---|--------------------|-------|-------|
|   | L2-CE              | HM    | BM    |
| 1. Osteopathic Principles, Practice, and Manipulative Treatment                   | 10.0%              | 0.0%  | 15.0% |
| 2. Osteopathic Patient Care and Procedural Skills                                 | 30.0%              | 0.0%  | 25.0% |
| 3. Application of Knowledge for Osteopathic Medical Practice                      | 26.0%              | 0.0%  | 15.0% |
| 4. Practice-Based Learning and Improvement in Osteopathic Medical Practice        | 7.0%               | 0.0%  | 5.0%  |
| 5. Interpersonal and Communication Skills in the Practice of Osteopathic Medicine | 5.0%               | 60.0% | 20.0% |
| 6. Professionalism in the Practice of Osteopathic Medicine                        | 7.0%               | 30.0% | 6.0%  |
| 7. Systems Based Practice in Osteopathic Medicine                                 | 5.0%               | 0.0%  | 5.0%  |

COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination of the United States of America; L2-CE, Level 2-Cognitive Evaluation; L2-PE, Level 2-Performance Evaluation; HM, Humanistic Domain; BM, Biomedical/Biomechanical Domain.

medicine and encompassed 97.5% of the 6,806 candidates who took the L2-CE during the complete 2019–2020 test cycle. The scores analyzed for this study were total score for the L2-CE, L2-CE scores for CD1 through CD7, and L2-PE scores for the HM and BM domains. All scores were on continuous scales.

Pearson and disattenuated correlations were calculated; disattenuated correlations were calculated to correct for measurement error [8, 9]. Disattenuated correlations have been used in similar previous studies [7, 10]. Cronbach's alpha [11] was used as the reliability estimate for L2-CE and CD1 through CD7. Generalizability coefficients [12] were used as reliability estimates for HM and BM.

## Results

Table 2 shows the Pearson correlations, disattenuated correlations, and reliability results for this study.

Pearson correlations ranged from 0.10 to 0.88 and were all statically significant ( $p < 0.01$ ). L2-CE total score was most strongly correlated with CD2 (0.88) and CD3 (0.85). Pearson correlations between the L2-CE competency domain subscores ranged from 0.17 to 0.70, and correlations which included either HM or BM ranged from 0.10 to 0.34, with the strongest of those correlations between BM and L2-CE total score (0.34) and between HM and BM (0.28).

As expected, disattenuated correlations were larger than the corresponding Pearson correlations and ranged from 0.18 to 1.00. The largest increase between corresponding Pearson and disattenuated correlations was for pairs of scores with lower reliabilities such as CD5 and

CD6, which had a Pearson correlation of 0.17 and a disattenuated correlation of 0.68. The smallest increase in correlations was observed in pairs of scores with larger reliabilities, such as L2-CE total score and HM, which had a Pearson correlation of 0.23 and a disattenuated correlation of 0.28.

## Discussion

The reliabilities for L2-CE, HM, and BM in this study were similar to those observed in comparable examinations. The reliability of L2-CE (0.87) is considered acceptable for a high stakes examination [13]. The reliabilities for HM (0.81) and BM (0.73) were similar to what has been observed in other medical licensure performance based examinations [6, 10]. The reliabilities for the L2-CE competency domain scores, which ranged from 0.22 to 0.74, were notably lower due to being comprised of smaller numbers of items; however, those scores are not recommended for use in high stakes decision making. Only disattenuated correlations are discussed below due to these lower reliabilities.

The correlation coefficients from our study results generally presented as expected. There were small to moderate correlations between the Level 2-CE total score and the two domains of the Level 2-PE, HM and BM. This indicates that L2-CE and L2-PE measure related but separate constructs, which supports the expectation that these examinations are related because they share the same master blueprint, with different percentages of items assigned to

**Table 2:** Correlations between COMLEX-USA L2-CE, L2-CE Competency Domain subscores (CD1 through CD7), and COMLEX-USA L2-PE scores for the HM and BM.<sup>a</sup>

|       |             | L2-CE       |             |             |             |             |             |             | L2-PE       |             |             |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|       |             | Total score | CD1         | CD2         | CD3         | CD4         | CD5         | CD6         | CD7         | HM          | BM          |
| L2-CE | Total score | <b>0.87</b> | 0.88        | 1.00        | 1.00        | 0.98        | 0.96        | 0.76        | 0.96        | 0.28        | 0.43        |
|       | CD1         | 0.63        | <b>0.58</b> | 0.64        | 0.64        | 0.63        | 0.52        | 0.44        | 0.53        | 0.28        | 0.39        |
|       | CD2         | 0.88        | 0.42        | <b>0.74</b> | 0.96        | 0.79        | 0.74        | 0.50        | 0.72        | 0.23        | 0.41        |
|       | CD3         | 0.85        | 0.41        | 0.70        | <b>0.71</b> | 0.78        | 0.70        | 0.44        | 0.69        | 0.19        | 0.36        |
|       | CD4         | 0.60        | 0.31        | 0.45        | 0.43        | <b>0.43</b> | 0.73        | 0.50        | 0.69        | 0.18        | 0.33        |
|       | CD5         | 0.42        | 0.18        | 0.30        | 0.28        | 0.22        | <b>0.22</b> | 0.68        | 0.88        | 0.39        | 0.40        |
|       | CD6         | 0.38        | 0.18        | 0.23        | 0.20        | 0.18        | 0.17        | <b>0.30</b> | 0.77        | 0.38        | 0.31        |
| L2-PE | CD7         | 0.45        | 0.20        | 0.31        | 0.29        | 0.23        | 0.21        | 0.21        | <b>0.25</b> | 0.37        | 0.34        |
|       | HM          | 0.23        | 0.19        | 0.18        | 0.14        | 0.10        | 0.16        | 0.19        | 0.17        | <b>0.81</b> | 0.37        |
|       | BM          | 0.34        | 0.25        | 0.30        | 0.26        | 0.19        | 0.16        | 0.14        | 0.15        | 0.28        | <b>0.73</b> |

<sup>a</sup>Pearson correlations are below the diagonal. Reliabilities are on the diagonal and in bold. Disattenuated correlations are above the diagonal. Disattenuated correlations greater than 1.00 are reported as 1.00.  $p < 0.01$  level. COMLEX-USA, Comprehensive Osteopathic Medical Licensing Examination of the United States of America; L2-CE, Level 2-Cognitive Evaluation; L2-PE, Level 2-Performance Evaluation; HM, Humanistic Domain; BM, Biomedical/Biomechanical domain.

each competency domain for each examination, but are still sufficiently different to justify both examinations. Additionally, the design of the master blueprint is supported by the differences in the strengths of the correlations between L2-CE and the L2-PE domains. The correlation between L2-CE and BM was larger than the correlation between L2-CE and HM; this difference reflects the design of the master blueprint, which requires that L2-CE and BM measure skills from all seven competency domains (CD1 through CD7), while HM measures only competency domains five and six (CD5 and CD6).

The strength of the correlations between HM and L2-CE competency domain subscores as well as between BM and L2-CE competency domain subscores were similar to the percentage of items for each competency domain required by the L2-PE test specifications (Table 1) [2, 3]. Our results showed that the HM domain was moderately correlated with CD5 (Interpersonal and Communication Skills in the Practice of Osteopathic Medicine), CD6 (Professionalism in the Practice of Osteopathic Medicine), and CD7 (Systems-Based Practice in Osteopathic Medicine). HM had smaller correlations with the other competency domains. According to the test specifications, HM should be comprised of mostly CD5 and CD6. CD7 is not a requirement for HM, but CD7 is most strongly correlated with CD5 and CD6, so the correlation between CD7 and HM is not surprising.

BM had larger correlations with CD1, CD2, CD3, and CD5 than with CD4, CD6, or CD7. These differences in correlations correspond to the minimum percentage on the master blueprint (Table 1), such that the competency domains with larger percentages of items on BM had larger correlations than the competency domains with smaller percentages of items. The differences in correlations were not large, but that is unsurprising, since the competency domains are correlated with each other (Table 2). To clarify, skills related to all seven competency domains are required to perform well in the BM domain. BM purports to assess the student's ability to complete a history and perform a physical examination, to perform osteopathic manipulative treatment, and to document in a subjective, objective, assessment, and plan (SOAP) note format. In terms of specific competencies, these skills require knowledge of and ability to correctly perform osteopathic manipulative treatment [5] (CD1), to complete a focused history and physical examination (CD2), to have knowledge and apply it to the case at hand (CD3), to be able to communicate to obtain the correct information (CD5), and to document the patient encounter completely for the record, which arguably requires skills in all seven competency domains [4, 14–16].

## Limitations

Although this study provides clear concurrent validity evidence supporting the intended uses of L2-CE and L2-PE, validity of any measurement must be established through ongoing evaluation of related evidence [17]. Therefore, the findings in this study should be evaluated in combination with past and future research supporting the validity of L2-CE and L2-PE. Additionally, because the results of this study were based on data from 97.5% of candidates who completed the L2-CE during the complete 2019–2020 test cycle, we expect that these results are generalizable to the overall population of L2-CE test takers, with the limitation being that the data are from a single L2-CE test cycle.

## Conclusions

There are two conclusions to be drawn from this study. First, the validity of L2-PE is supported by the small to moderate correlations found with L2-CE in this study. The results support the use of both multiple choice and performance examinations to ensure the assessment of a broader range of competencies in osteopathic medicine. Second, the strength of the correlations between HM, BM, and the seven L2-CE competency domain subscores was generally reflective of the minimum percentage of items for each competency domain measured by HM and BM, as defined by the master blueprint. In other words, scores from HM and BM tended to be more strongly correlated with the L2-CE competency domain subscores in competency domains where the master blueprint specified larger percentages of items. This finding supports the concurrent validity of L2-CE and L2-PE. Overall, we believe this analysis supports the need for, validity of, and continued use of the L2-PE and L2-CE examinations.

**Research funding:** None reported.

**Author contributions:** All authors provided substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; all authors drafted the article or revised it critically for important intellectual content; all authors gave final approval of the version of the article to be published; and all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Competing interests:** Drs. Craig, Wang, Tsai, and Sandella are employees of the National Board of Osteopathic

Medical Examiners and therefore have a financial stake in the success of the COMLEX-USA.

## References

1. National Board of Osteopathic Medical Examiners. COMLEX-USA bulletin of information 2020-2021. Available from: <https://www.nbome.org/exams-assessments/comlex-usa/bulletin/> [Accessed 16 October 2020].
2. COMLEX-USA Level 2-CE. National Board of Osteopathic Medical Examiners website. Available from: <https://www.nbome.org/exams-assessments/comlex-usa/comlex-usa-level-2-ce/> [Accessed on 16 October 2020].
3. COMLEX-USA Level 2-PE. National Board of Osteopathic Medical Examiners website. Available from: <https://www.nbome.org/exams-assessments/comlex-usa/comlex-usa-level-2-pe/> [Accessed on 16 October 2020].
4. Gimpel JR, Boulet DO, Errichetti AM. Evaluating the clinical skills of osteopathic medical students. *J Am Osteopath Assoc* 2003;103:267–79.
5. Boulet JR, Gimpel JR, Dowling DJ, Finley M. Assessing the ability of medical students to perform osteopathic manipulative treatment techniques. *J Am Osteopath Assoc* 2004;104:203–11.
6. Baker HH, Cope MK, Adelman MD, Schuler S, Foster RW, Gimpel JR. Relationships between scores on the COMLEX-USA Level 2-Performance Evaluation and selected school-based performance measures. *J Am Osteopath Assoc* 2006;106:290–5.
7. O'Neill TR, Peabody MR, Song H. The predictive validity of the National Board of Osteopathic Medical Examiners' COMLEX-USA examinations with regard to outcomes on American Board of Family Medicine Examinations. *Acad Med* 2016;91:1568–75.
8. Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;15:72.
9. Schmidt FL, Hunter JE. Measurement error in psychological research: lessons from 26 research scenarios. *Psychol Methods* 1996;1:199–223.
10. Harik P, Clauser BE, Grabovsky I, Margolis MJ, Dillon GF, Boulet JR. Relationships among subcomponents of the USMLE Step 2 Clinical Skills Examination, the Step 1, and the Step 2 Clinical Knowledge Examinations. *Acad Med* 2006;81(10 Suppl):S21–4.
11. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
12. Brennan RL. Generalizability theory. New York, NY: Springer-Verlag; 2001.
13. Webb NW, Shavelson RJ, Haertel EH. Reliability coefficient and generalizability theory. In: Rao CR, Sinharay RS, editors. *Handbook of statistics*, vol 26. Elsevier; 2006:81–124 pp.
14. Roberts WL, Solomon M, Langenau E. An investigation of construct validity of humanistic clinical skills on a medical licensure examination. *Patient Educ Counsel* 2011;82:214–21.
15. Sandella JM, Roberts WL, Gallagher LA, Gimpel JR, Langenau EE, Boulet JR. Patient note fabrication and consequences of unprofessional behavior in a high-stakes clinical skills licensing examination. *Acad Med* 2009;84(10 Suppl):S70–3.
16. Sandella JM, Boulet JR, Langenau EE. An evaluation of cost and appropriateness of care as recommended by candidates on a national clinical skills examination. *Teach Learn Med* 2012;24:303–8.
17. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *ETS Res Rep Ser* 1994;1994:i–28.