Equitable Exposure, Unequal Outcomes

Summative Performance Following Clinical Integration Anatomy Laboratories

Christopher Ramus, BS, MS-2[1], Stefanie Attardi, PhD[2]

[1]M.D. Candidate, Oakland University William Beaumont School of Medicine

[2]Associate Professor, Department of Foundational Medical Studies, Oakland University William Beaumont School of Medicine

Background

- Cadaver-based anatomy education remains essential but faces high content demands, limited physical and personnel resources, and a continually condensed timeline.[1,2]
- These challenges necessitate efficient instructional efficient instructional strategies. Pedagogical research supports active learning strategies for better retention, engagement, and application.[3]

Clinical Integration Laboratories at OUWB:

- CILs in M1 Anatomical Foundations of Clinical Practice (AFCP) 1 and 2 courses were designed to help students use active learning to clinically apply dissection-based anatomy and prepare for summative assessment.
- Small teams (5-6 students) progressed through six case-based stations (Figure 1) that followed the same format:
  - 1 min: Individually answer 1 vignette-based multiple choice question (MCQ)
  - 6 mins: Student-led team discussion about the MCQ
  - 7 mins: Team identification of relevant cadaveric structures

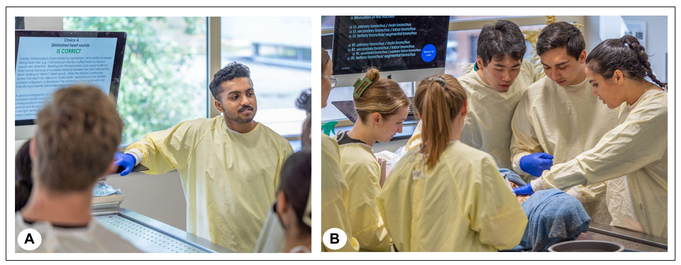

Figure 1. Typical Clinical Integration Laboratory station
Each station had a computer monitor for displaying the question, answer choice rationales, and structure list, as well as 1-2 dissected body donors. **(A)** Student team discussing answers to the vignette-based question. **(B)** Student team identifying relevant cadaveric structures.
Photography courtesy of Dr. Malli Barremkala.

- Students responded favorably to CILs on course evaluations; **however, a quantitative relationship between CIL and summative exam performance has not been established.**

Purpose & Research Objectives

Determine whether CIL performance correlates with summative exam performance and whether prior AFCP achievement influences CIL and summative exam performance.

Objective 1: Determine whether CIL formative assessment scores in AFCP2 correlate with:

a. Written examination score (WEX)
b. Written examination sub-scores mapped to CIL cases (WEX Sub)
c. Practical examination scores (PEX)
d. Practical examination sub-scores mapped to CIL cases (PEX Sub)

Objective 2: Determine whether AFCP1 performance quartiles differ in terms of AFCP2 measures:

b. CIL score
c. Written examination score
d. Written examination sub-scores mapped to CIL cases
e. Practical examination scores
f. Practical examination sub-scores mapped to CIL cases

Methods

- This study was determined by Oakland University IRB to be *no human subjects research* (IRB-FY2026-42).
- M1 students (N=128) in the 2025 AFCP2 course completed two CILs and two summative exams. The data collection timeline with study variables is presented in Figure 2.

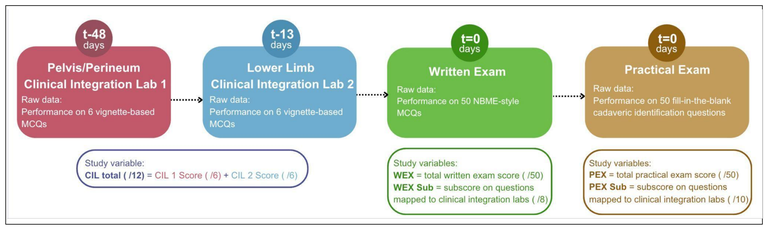

Figure 2. Data Collection and Study Variables
Data were collected during CILs using pencils and paper, and examinations were administered using Examplify (Examsoft, Boca Raton, Florida).
t: time until summative exams; MCQ: multiple choice question; NBME: National Board of Medical Examiners.

- Statistical analyses were performed using SPSS version 31 (IBM, Armonk, NY)
- **Addressing Objective 1:** Pearson correlations examined associations between CIL scores and WEX, WEX Sub, PEX, and PEX Sub scores. α=0.05.
- **Addressing Objective 2:** Performance quartiles were determined using total course scores from the previous anatomy course (AFCP1). Kruskal-wallis tests and Dunn's post hoc analyses with Bonferroni correction compared the CIL, WEX, WEX Sub, PEX, and PEX Sub scores across performance quartiles. α=0.05.

Results

- **Addressing Objective 1:** CIL total scores had small but significant (p=0.003) positive correlations with WEX score (r=0.261) and WEX subscore (r=0.258), as well as total PEX score (r=0.188, p=0.035).
- **Addressing Objective 2:**
  - Significant differences were observed across quartiles for the WEX (H(3)=64.227,p<0.001), WEX Sub (H(3)=29.186,p<0.01), PEX (H(3)=54.960,p<0.001), and PEX Sub (H(3)=32.548,p<0.001).
  - Dunn's post hoc analyses with Bonferroni correction for scores related to the Written Exam and Practical Exam are presented in Figures 3 and 4, respectively.
- All other analyses were not significant (p>0.05).

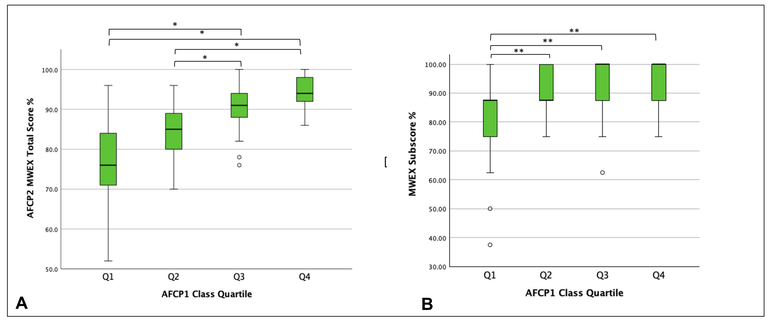

Figure 3. Written Examination Score (WEX) and Written Examination Subscore (WEX Sub) by Quartile
Dunn's post hoc tests with Bonferroni correction for the Independent samples Kruskal-Wallis tests (n = 128). **(A)** Q4 and Q3 each scored significantly higher than Q2 and Q1 on the WEX (*p≤0.020). **(B)** Q4, Q3, and Q2 each scored significantly higher than Q1 on the WEX Sub (**p≤0.025).
Q: Quartile (ranges 1-4, where Q4 is the top quartile and Q1 is the bottom quartile).

Results (continued)

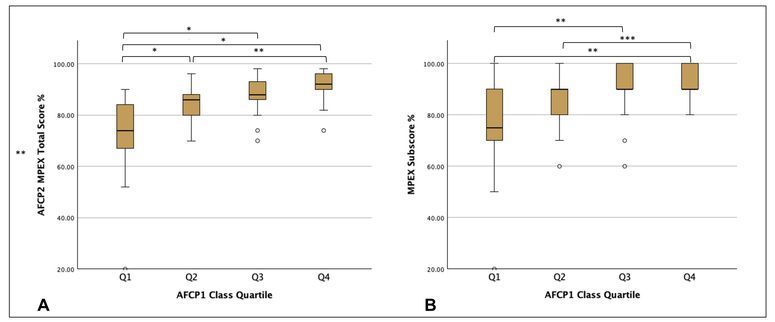

Figure 4. Practical Exam Score (PEX) and Practical Exam Subscore (PEX Sub) by Quartile
Dunn's post hoc tests with Bonferroni correction for the Independent samples Kruskal-Wallis tests (n = 128). **(A)** Q4, Q3, and Q2 each scored significantly higher than Q1 on the PEX (*p≤0.022) and Q4 scored significantly higher than Q2 (**p<0.001). **(B)** Q4 and Q3 scored significantly higher than Q1 (**p<0.001). Q4 scored significantly higher than Q2 (***p=0.043)
Q: Quartile (ranges 1-4, where Q4 is the top quartile and Q1 is the bottom quartile).

Discussion

- Students performed similarly on CIL formative assessments, but the bottom quartile scored significantly lower on summative exams, even on CIL-related questions.
- Team-based formative activities did not fully predict individual exam performance.
- Consistent with medical education literature, summative assessments more strongly influence student study behaviors compared to formative assessments, which may explain why performance quartiles differed during summative exams.[4]

Future Directions:

- Analyze Team-Level Effects: Assess whether individual performance is influenced by specific Clinical Integration Laboratory (CIL) team assignments.
- Expand Data Collection: Analyze additional CIL units with varied anatomical content to assess whether performance patterns differ across topics.

Conclusion

While CILs promote statistically equivalent formative performance, this did not translate to summative exam performance for students in the lowest quartile. The results highlight a critical limitation of aggregate instructional design in anatomy education: equal exposure does not ensure equitable outcomes, underscoring the need for targeted support for knowledge transfer in lower-performing students.

References

1. Drake RL, McBride JM, Pawlina W. An update on the status of anatomical sciences education in United States medical schools. Anat Sci Educ. 2014;7(4):321–325. doi: 10.1002/ase.1468
2. McBride JM, Drake RL. National survey on anatomical sciences in medical education. Anat Sci Educ. 2018;11(1):7–14. doi: 10.1002/ase.1760
3. Bin Abdulrahman KA, Khalaf AM, Bin Abbas FB, Alanazi OT. Study habits of highly effective medical students. Adv Med Educ Pract. 2021;12:627–633. doi: 10.2147/AMEP.S309535
4. Raupach T, Brown J, Anders S, Hasenfuss G, Harendza S. Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. BMC Medicine. 2013;11(1). doi:https://doi.org/10.1186/1741-7015-11-61

Acknowledgments

The authors wish to thank the AFCP course directors, Drs. Jickssa Gemechu, Varna Taranikanti, and Ngan Nguyen for research consultation and assistance with obtaining the study data, and Dr. Malli Barremkala and Dan Schlegel for assistance with logistics planning in the laboratory. We also wish to thank the individuals who have donated their bodies to medical education.