

## RESEARCH LETTER

# Novel quality assessment methodology in focused cardiac ultrasound

Cardiac point-of-care ultrasound (POCUS) has become a fundamental component of the evaluation of patients in the emergency department (ED) to diagnose cardiac pathology.<sup>1</sup> Quality assessment (QA) is one of the six required elements of diagnostic POCUS examinations per the American College of Emergency Physicians (ACEP).<sup>2</sup> QA is routinely performed to ensure standard of patient care is met and to assess for competency, particularly at the trainee level. It has recently been described that 82% of EM residency programs report use of QA as an assessment tool.<sup>3</sup> It is imperative to provide a reliable scoring system to limit inconsistencies in the way we are measuring clinical skill as it relates to ultrasound.

The current grading scale used for QA that is endorsed by ACEP was developed from a consensus report of emergency ultrasound leaders indicating a need for a systematic method to report and communicate POCUS findings.<sup>2</sup> The ACEP grading scale is a nonspecific grading classification that applies regardless of the type of study performed.<sup>2</sup> This contrasts with other QA grading systems that have been described in an organ-specific manner. Examples of these systems include focused cardiac ultrasound assessment demonstrated by Kimura et al.<sup>4</sup> and focused gynecological emergency ultrasound examination by Salomon.<sup>5</sup> The goal of this study was to determine whether a similar, organ-specific grading scale would be a more reliable method of assessment with improved interobserver agreement. Furthermore, we sought to determine whether an organ-specific grading scale had more variance in the scores that are chosen and therefore more focused feedback to the sonographers performing the studies.

We conducted a prospective analysis of the first 200 cardiac POCUS studies performed in our ED that were submitted for QA into our image database in the year of 2020. Four reviewers, who were either emergency ultrasound fellowship trained or current emergency ultrasound fellows with at least 9 months of QA experience, scored each of the studies. Two reviewers used the current ACEP grading scale: 1 = no recognizable structures; 2 = minimally recognizable structures but insufficient for diagnosis; 3 = minimal criteria met for diagnosis, recognizable structures but with some technical or other flaws; 4 = minimal criteria met for diagnosis, all structures imaged well; and 5 = minimal criteria met for diagnosis, all structures

imaged with excellent image quality.<sup>2</sup> The two other reviewers used a cardiac-specific grading scale as previous described by Kimura et al.: 0 = no image obtained; 1 = only cardiac motion detected; 2 = chambers and valves grossly resolved; 3 = endocardium and wall thickness seen but incomplete; and 4 = greater than 90% of endocardium and valve motion seen.<sup>3</sup> The primary outcomes were the level of agreement between the reviewers, indicating the reliability of the scoring system, and the variability of the scores given to the studies. This study was approved by the VCU institutional review board.

The ACEP score was on a scale of 1–5, while the organ-specific score was on a scale of 0–4. For equal comparison, we added 1 to the organ-specific grading. For the primary outcome, the intraclass correlation coefficient (ICC) based on two-way random-effect model with a single rater for each grading scale was computed. Ten thousand bootstrapped ICCs were generated to construct 95% confidence intervals (CIs) for both grading systems, and a two-sided one-sample t-test was used to determine if there were differences in the bootstrapped ICCs between the two grading systems. The ICC between reviewers for the ACEP grading scale was 0.54 (95% CI 0.410–0.555) indicating moderate agreement, while the ICC between reviewers using the organ-specific grading scale was 0.75 (95% CI 0.600–0.769) indicating good agreement. These findings were statistically significant with one-sample t-test p-value of <0.0001. The ACEP grading scale mean ( $\pm$ SD) was 3.15 ( $\pm$ 0.693) versus 4.16 ( $\pm$ 0.967) for the organ-specific grading system. A 95% CI for the variance ratio was constructed to determine whether there were differences in the variability between the two grading systems. The variance of scores using the organ-specific grading scale was found to be more than 1.95 times greater when compared to the scores using ACEP grading scale. The variance among each group of scores using a 95% CI of the bootstrapped variance ratio (1.49–2.51), the organ-specific grading scale was found to have significantly more variability than the ACEP scoring system. The summary figure for the ACEP and organ-specific grading scale are found in [Figure 1](#).

This study is a quantitative prospective analysis comparing the current ACEP recommended QA method to an organ-specific method. We found that there was an increased interobserver agreement between reviewers and increased variability of the scores given to the studies when using the organ-specific method compared to the ACEP method. This suggests that added detail in

Presented at ACEP Research Forum, October 2021.

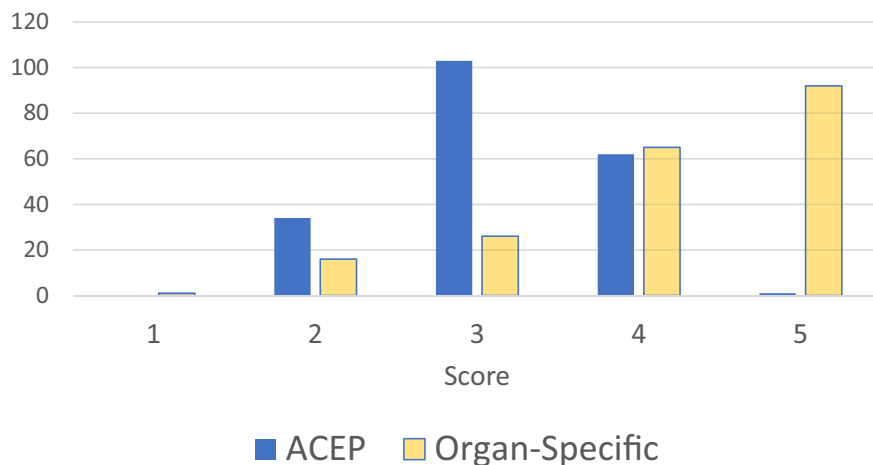
Supervising Editor: Dr. Daniel Theodoro.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Academic Emergency Medicine* published by Wiley Periodicals LLC on behalf of Society for Academic Emergency Medicine.

## Scores by Grading Scale

**FIGURE 1** Scores given to studies by grading scale



group	min	Q1	median	Q3	max	mean	sd	n	missing
Organ-Specific	1	4	4	5	5	4.155	0.9672917	200	0
ACEP	2	3	3	4	5	3.150	0.6927478	200	0

the guidance to the reviewers led to increased use of the spectrum of the scores. This is particularly important for trainees; they are often the performers of the ultrasound and, as such, the recipient of the QA reports. More direct feedback may expand their knowledge base to be implemented in future studies. For example, the top score on the ACEP grading scale states in a nonspecific manner, “all structures imaged with excellent image quality” versus the cardiac organ-specific grading scale states “greater than 90% of endocardium and valve motion seen.” This allows the learner to understand why the top score was given and the structures needed to meet this requirement. The opposite is also true, the organ-specific grading scale presents a detailed response and as such suggests direction for improvement on subsequent studies. The use of a standardized focused scoring system may also improve the quality of the images themselves. Salomon<sup>4</sup> demonstrated that the implementation of an organ specific gynecological emergency ultrasound may improve the quality level of ultrasound examinations performed.

Our data highlight the importance of QA methodology that promotes objectivity demonstrated by high interobserver agreement. POCUS examinations in the ED continue to rise, and as they do there is a need for a standardized method to perform QA that has been thoroughly validated prior to implementation. This single-center study is limited by the images being obtained and reviewed at an academic ED by ultrasound-trained faculty limiting the generalizability.

This investigation demonstrates an organ-specific quality assurance of focused cardiac ultrasound procedures performed in the ED may give the providers performing the studies a wider range of feedback with increased inter-rater agreement among those doing the reviews. Given these findings, there may be a benefit to moving toward an organ-specific QA scale for POCUS studies obtained in the ED.

### KEYWORDS

cardiac point of care ultrasound, cardiac ultrasound, emergency department, quality assessment

### ACKNOWLEDGMENT

The authors thank Chen Wang and Martin Lavalley with the VCU Biostatistical Consulting Laboratory, which is partially supported by award UL1R002649 from the National Institutes of Health's National Center for Advancing Translational Science, Department of Biostatistics, VCU School of Medicine.

Stephanie DeMasi MD<sup>1</sup> 

Lindsay A. Taylor MD<sup>1</sup> 

Adam Weltler MD<sup>1</sup>

John C. Wiggins MD<sup>1</sup>

Jacob Wayman MD<sup>1</sup>

Chen Wang MS<sup>2</sup>

David P. Evans MD<sup>1</sup>

Jessica R. Balderston MD<sup>1</sup> 

<sup>1</sup>Department of Emergency Medicine, Virginia Commonwealth University, Richmond, Virginia, USA

<sup>2</sup>Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia, USA

### Correspondence

Jessica R. Balderston, MD, Department of Emergency Medicine, Virginia Commonwealth University, Richmond, VA 23298, USA.

Email: [jessica.balderston@vcuhealth.org](mailto:jessica.balderston@vcuhealth.org)

**ORCID**

Stephanie DeMasi  <https://orcid.org/0000-0002-4149-8955>

Lindsay A. Taylor  <https://orcid.org/0000-0001-7397-8665>

Jessica R. Balderston  <https://orcid.org/0000-0002-5315-3604>

**REFERENCES**

1. Pattock AM, Kim MM, Kersey CB, et al. Cardiac point-of-care ultrasound publication trends. *Echocardiography*. 2022;39(2):240-247.
2. Liu RB, Blaivas M, Moore C, Sivitz AB, Flannigan M, Tirado A., et. al. *Emergency Ultrasound Standard Reporting Guidelines*. ACEP Emergency Ultrasound Guidelines. June 2018.
3. Amini R, Adhikari S, Fiorello A. Ultrasound competency assessment in emergency medicine residency programs. *Acad Emerg Med*. 2014;21:799-801.
4. Kimura BJ, Gilcrease GW 3rd, Showalter BK, Phan JN, Wolfson T. Diagnostic performance of a pocket-sized ultrasound device for quick-look cardiac imaging. *Am J Emerg Med*. 2012;30(1):32-36.
5. Salomon LJ. A Score-based Method to Improve the Quality of Emergency Gynaecological Ultrasound Examinations. *Eur J Obstet Gynecol Reprod Biol*. 2009;143(2):116-120.