RESEARCH

Tool to improve qualitative assessment of left ventricular systolic function

Daniel P Walsh MD, Kadhiresan R Murugappan MD, Achikam Oren-Grinberg MD MS, Vanessa T Wong BS, John D Mitchell MD and Robina Matyal MD

Department of Anesthesia, Critical Care and Pain Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA

Correspondence should be addressed to D P Walsh or V T Wong: dpwalsh@bidmc.harvard.edu or vtwong@bidmc.harvard.edu

Abstract

Interactive online learning tools have revolutionized graduate medical education and can impart echocardiographic image interpretive skills. We created self-paced, interactive online training modules using a repository of echocardiography videos of normal and various degrees of abnormal left ventricles. In this study, we tested the feasibility of this learning tool. Thirteen anesthesia interns took a pre-test and then had 3 weeks to complete the training modules on their own time before taking a post-test. The average score on the post-test (74.6% ± 11.08%) was higher than the average score on the pre-test (57.7% ± 9.27%) (P < 0.001). Scores did not differ between extreme function (severe dysfunction or hyperdynamic function) and non-extreme function (normal function or mild or moderate dysfunction) questions on both the pre-test (P = 0.278) and post-test (P = 0.093). The interns scored higher on the post-test than the pre-test on both extreme (P = 0.0062) and non-extreme (P = 0.0083) questions. After using an online educational tool that allowed learning on their own time and pace, trainees improved their ability to correctly categorize left ventricular systolic function. Left ventricular systolic function is often a key echocardiographic question that can be difficult to master. The promising performance of this educational resource may lead to more time- and cost-effective methods for improving diagnostic accuracy among learners.

Introduction

As the use of echocardiography expands in perioperative and critical care settings, achievement of proficiency in this technology is becoming a standard of care in multiple clinical settings (https://www.acgme.org/Portals/0/PDFs/Milestones/AnesthesiologyMilestones.pdf and https://www.echoboards.org/EchoBoards/News/2019_Adult_Critical_Care_Echocardiography_Exam.aspx) (1, 2). Multiple programs have introduced the concept of ‘pre-clinical proficiency’ in which the trainees are required to demonstrate a certain level of understanding of the basic concepts and clinical workflow prior to patient exposure (3, 4, 5). Specifically, for echocardiography, mixed haptic simulators have been successfully used for imparting a level of basic pre-clinical proficiency (6, 7, 8, 9). Whereas these are invaluable tools for acquisition of psychomotor skills, learning of echocardiographic image interpretation requires trainee presence in the clinical environment or physical interaction with the equipment. Specifically, the ability to appreciate gross and subtle changes in ventricular function is acquired through repetitive visual observations and pattern recognition of various clinical examples in the setting of expert feedback (10).
Availability of interactive online learning tools has further revolutionized the graduate medical education and testing, allowing interactive and continuous access through smart communication devices without the constraints of time and space. Online availability of these educational materials is specifically an invaluable resource for imparting the echocardiographic image interpretive skills that are based on repetitive visual exposure to a diverse range of normal and pathologic exams. Using a commercially available learning management software application, we created a self-paced, self-testing, and interactive online repository of echocardiography media clips of normal and various degrees of abnormal left ventricles (LV) for our residents. By creating this resource to be used online, potentially by hand-held smart devices, we seek to circumvent some of the aforementioned obstacles of adding an increased didactic burden onto the clinically busy and work hour constrained schedule of medical trainees. In this study, we tested the feasibility of using this online educational resource to teach basic echocardiography image interpretation skills to anesthesia residents.

Materials and methods

This study received institutional review board approval for exempt status by the Committee on Clinical Investigations at Beth Israel Deaconess Medical Center (BIDMC).

Development of the training program

A cache was created of de-identified video clips from transthoracic parasternal long axis, parasternal short axis, apical four-chamber, and subcostal views of the LV in varying degrees of function. These views were used because they are the basic views that are frequently taught for point-of-care echocardiography. The apical two-chamber view could have been helpful to include, but only the most basic views were used, as this educational tool was geared to novice (postgraduate year 1) anesthesia trainees.

Video clips were obtained from examinations performed by registered cardiac sonographers. The five categories of function were hyperdynamic (defined as LV ejection fraction (EF) >75%), normal (LV EF 50–75%), mild systolic dysfunction (LV EF 40–50%), moderately reduced systolic dysfunction (LV EF 30–40%), and severely reduced dysfunction (LV EF <30%). While the official quantitative normal LV EF is 52% for men and 45% for women, an LV EF greater than 50% was chosen for normal function, as qualitative function does not provide the precision to differentiate between single digit percent levels of function (11). The true level of function of each clip was assessed by a formal interpretation by the cardiology echocardiography lab (12 readers certified in adult echocardiography by the National Board of Echocardiography (NBE)) using a combination of qualitative and quantitative methods as part of the standard interpretation process of echocardiographic exams at BIDMC. The combination varied from exam to exam, but Simpson's biplane method was the standard quantitative method for all exams. Some were assessed with 3D measurements as well.

After assessment by the cardiology echocardiography lab, the clips were reviewed by the primary author (DW) for image quality and to assess if the qualitative assessment seemed easily congruent. Multiple other authors (KM and AO) reviewed the clips as well. Clips were not entered into the cache if there was either disagreement between the qualitative and quantitative assessments on the official read or if they appeared to the authors to be too close to the border of two function level categories, thus having potential to confuse non-expert level trainees. In addition, clips that were believed to be of too poor image quality were not included in the cache. For exams with regional discrepancies, only views that consistently showed the regional abnormality were used; clips with extreme regional discrepancies were not used. The final cache consisted of several hundred video clips split across the four different echocardiographic windows listed, which were taken from 97 different echocardiographic exams across the five categories of function (15–24 different exams for each category of function).

Training modules were developed with Articulate Storyline authoring tool (Articulate Global Inc., New York, NY, USA), an interactive and online training system. Twenty-question pre- and post-tests were also developed to assess trainees’ ability to identify the varying categories of left ventricular systolic function before and after the training modules. The pre- and post-test questions were identical to prevent differences in clips from being a confounder in our comparison of performance on the two tests. In addition, no immediate feedback was given on the pre-test and post-test to prevent ‘learning’ from the test. They were designed by the primary author (DW) using clips reviewed by a study member (KM). Two study members (KM and AO) completed the pre/post-test to ensure quality prior to administration to the residents. The test consisted of twenty questions in order to have a power of >0.8 to detect a difference in means between 0.6 and 0.8 based on a power analysis with $\alpha=0.05$. 

https://erp.bioscientifica.com
https://doi.org/10.1530/ERP-19-0053
© 2020 The authors Published by Bioscientifica Ltd

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
The pre-test was taken before using any of the educational modules. A brief didactic presentation was made to instruct trainees on what to look for when qualitatively assessing LV EF. The core training module consisted of four training quizzes composed of ten video-based questions as a tool for trainees to calibrate their eyes to echocardiographic video clips. None of the video clips in the pre- or post-test were used in any of the training quizzes. Each question in the training quizzes had a subsequent feedback slide where multiple video clips with different categories of LV systolic function were played simultaneously side by side so that the trainee could ‘calibrate the eye’ to differentiate between the different categories of function. Repetition of video clips across the training quizzes was minimized, and clips that were used for comparison were kept from being used as question clips in order to maximize exposure to a variety of images and to avoid recognition of a particular image as opposed to calibrating the eye to the particular function level. A sample of the training quizzes can be viewed at https://anesthesiaeducation.net/moodle/course/view.php?id=185 (username: sample; password: sample). Varying degrees of image quality were included in the training modules to present trainees with realistic clinical scenarios, particularly in settings where optimal image quality may be limited by patient habitus, comorbidities, or other factors. However, the clips used for the pre- and post-test were clear with the best image quality to avoid confounding the ability to assess function.

Thirteen anesthesia residents in their internship year (postgraduate year 1) participated in a perioperative ultrasound course (including ultrasonic physics, knobology, and basic transthoracic echocardiography imaging techniques) as part of their regular didactics. They took the pre-test quiz and then had 3 weeks to complete the training modules online on their own time prior to taking the post-test. The pre- and post-tests and training modules were administered through the department’s online learning management system (LMS) (Moodle, Moodle Pty Ltd, West Perth, Australia). The LMS allowed tracking of completion of the quizzes, time taken to complete each quiz, and scores for each quiz. As part of research, they were asked to allow the study staff to use their data to assess the effectiveness of the training program. The residents were informed about the research aspect via email by the study coordinator and notified the study coordinator individually about their decision to accept or decline participation in the research. A sample size of 13 provided a power of 0.88 to detect a difference in means between 0.6 and 0.8 based on a power analysis with $\alpha=0.05$.

### Statistical analysis

Stata/Special Edition 12.1 (StataCorp LP, College Station, TX, USA) was used for all analyses. A $P$-value of $<0.05$ was considered significant.

### Overall performance

Pre- and post-test scores were recorded, averaged, and compared using a paired two-tailed $t$-test.

In order to minimize any selection bias from the pre- and post-test design process, we administered the test to five attending anesthesiologists who specialize in cardiac anesthesia and/or critical care (three were certified by the NBE and two were testamurs of the NBE) and who were not involved in the design of the test. These attendings were classified as ‘experts’ in the study. We compared the interns’ pre- and post-test scores to the scores of the experts using an unpaired, two-sample, two-tailed $t$-test as a way to assess intern performance against a standard, acceptable skill level for attending anesthesiologists specialized in cardiac anesthesia and/or critical care.

### Performance by function level

Questions on the pre- and post-test were separated into two categories: (1) questions with ‘extreme’ functions (severe dysfunction or hyperdynamic function) and (2) questions with ‘non-extreme’ functions (normal function, mild dysfunction, or moderate dysfunction). Since it is likely easier to identify severe dysfunction or hyperdynamic function than to differentiate the less extreme functions, we wanted to determine if any improvement we detected in the primary analysis was driven by questions with ‘extreme’ functions. We therefore conducted a secondary analysis to determine if function level had an association with the scores. Wilcoxon signed-rank tests were used to compare performance between the dichotomized artificial function categories (‘extreme’ and ‘non-extreme’). The median score and interquartile range (IQR) were calculated for each artificial function category on each (pre and post) test.

### Results

#### Overall performance

All 13 trainees completed the educational program. The average time to complete a training quiz was 11.2 min, with less than an hour of total time to complete all four training quiz modules. Average test results are shown in Fig. 1A; individual test results of the 13 trainees are shown in Fig. 1B. The average score on the pre-test was...
57.7%±9.27% correct. After using the educational tool, the average score on the post-test was 74.6%±11.08% correct, which was a statistically significant increase from the pre-test (P<0.001). The residents scored the same on the post-test as they did on the pre-test.

In comparison, the average score of the experts (70%±10.61%) was significantly higher than the interns’ average pre-test score (P=0.027) and not significantly different than the interns’ average post-test score (P=0.435). On the post-test, all the interns scored within one s.d. of the experts’ average score or higher.

**Performance by function level**

Overall, in both the pre- and post-tests, the interns appeared to score higher on the ‘extreme function category’ (severe dysfunction and hyperdynamic function) questions (median: 71; IQR: 57 to 86) than on the ‘non-extreme category’ (normal function and mild or moderate dysfunction) questions (median: 62; IQR: 54 to 69), but this difference was not significant (P=0.057). On the pre-test, there was no significant difference between the extreme (median: 57; IQR: 57 to 71) and non-extreme (median: 62, IQR: 54 to 62) questions (P=0.278). On the post-test, there was no significant difference between the extreme (median: 86; IQR: 71 to 86) and non-extreme (median: 69, IQR: 54 to 85) questions (P=0.093).

The interns scored significantly higher on the post-test (median: 86; IQR: 71 to 86) than the pre-test (median: 57; IQR: 57 to 71) on the extreme questions (P=0.0062). They also scored significantly higher on the post-test (median: 69, IQR: 54 to 85) than the pre-test (median: 62, IQR: 54 to 62) on the non-extreme questions (P=0.0083). Figure 2 shows the median and interquartile ranges for the pre- and post-tests based on extreme or non-extreme function.

**Discussion**

This was a pilot educational intervention with the goal to improve cardiac ultrasound image interpretation skill by trainees, specifically, qualitative interpretation of LV systolic function. Many existing educational echocardiography tools that we have come across in our review of the literature focus on didactics related to facts about echocardiography, some of which may be about how to assess LV function, and have assessments that tested more basic knowledge of echocardiographic facts (6, 12, 13). This is different than a focus on developing and applying specifically the skill of image interpretation as we aimed to do here. Based on Donald Kirkpatrick’s Four Level Training Evaluation Model (https://educationatechnology.net/kirkpatrick-model-four-levels-learning-evaluation/), it is a higher level of understanding to be able to apply knowledge to assess LV function than to just demonstrate knowledge of factors related to wall motion assessment.
levels of function. Thus, it is possible to develop a competency based echocardiographic curriculum with cognitive, manual, and interpretation skill sets for the assessment of cardiac function. This methodology could also be expanded to develop interpretation skill in echocardiographic assessment of valvular and right ventricular function.

An issue with ultrasound training, in general, is that it can require significant resources, and an advantage of this approach is that experience be gained without coordinating the concurrent use of ultrasound equipment, patients with varying function levels, or an expert to give feedback with the trainee’s schedule. Trainees can learn to recognize normal and variations of abnormal function without a physical presence in the echo laboratory, having a fixed personal computer, or evaluating multiple patients in real-time. Another advantage of this approach is the side-by-side video comparison of different LV function levels to highlight the difference of function. It is one thing to look at a video clip of mild LV dysfunction in isolation and be told by an expert that it is mild dysfunction, and it is another to see simultaneous side by side clips of normal, mild, and moderate dysfunction, whereby the differences are immediately visible and the subtle continuum of normal function to dysfunction can be appreciated by the trainee. Additionally, this resource can be used online via a hand-held smart device if desired for availability at any time. Comprehensive exposure to the assessment of LV function is a difficult task that requires evaluation of a large number of patients. Indeed, even among cardiology trainees, the most frequent source of disagreement with attending overread is regarding LV function (19). This program took trainees less than 1 hour in total time and allowed trainees to be exposed to almost 100 exams of the LV in much less time than real-time examination of 100 patients by echocardiography.

It is interesting to note that three of the trainees did not improve their scores on the post-test. Similar to most of the other interns, they also did not have any formal experience or training with ultrasound before the course. This could be another strength of this training module in that it could quickly identify trainees who may require more rigorous or remedial training to improve this skill. The current teaching models for perioperative echocardiography are gradually ensuring that the trainees are well prepared with the understanding and basic application for echocardiography; however, there is a paucity of continuous training models to ensure retention of knowledge and competence to recognize

![Figure 2](https://erp.bioscientifica.com)

**Figure 2**
Median and interquartile ranges (IQR) for the pre- and post-test based on extreme or non-extreme function. On the pre-test, there was no significant difference between the extreme (median: 57; IQR: 57 to 71) and non-extreme (median: 62, IQR: 54 to 62) questions ($P = 0.278$ based on a Wilcoxon signed-rank test with $\alpha = 0.05$). Likewise, on the post-test, there was no significant difference between the extreme (median: 86; IQR: 71 to 86) and non-extreme (median: 69, IQR: 54 to 85) questions ($P = 0.093$ based on a Wilcoxon signed-rank test with $\alpha = 0.05$). The interns scored significantly higher on the post-test (median: 86; IQR: 71 to 86) than the pre-test (median: 57; IQR: 57 to 71) on the extreme questions ($P = 0.0062$ based on a Wilcoxon signed-rank test with $\alpha = 0.05$). They also scored significantly higher on the post-test (median: 69, IQR: 54 to 85) than the pre-test (median: 62, IQR: 54 to 62) on the non-extreme questions ($P = 0.0083$ based on a Wilcoxon signed-rank test with $\alpha = 0.05$). The medians are indicated by the red lines in the figure.

Other high-fidelity simulation-based educational programs have shown to be useful but often focus on improving skills such as image acquisition or anatomy identification and not necessarily on image interpretation (6, 8, 14). Image acquisition is an important skill, but again is different than the specific skill of applying knowledge and interpreting a range of images. The use of these simulators often involves complex scheduling with trainees’ clinical responsibilities as a barrier that online programs do not have. Training programs that we have come across that do test application and interpretation ability of trainees are lengthy programs that involve many hours of hands-on training alongside expert instructors (15, 16, 17, 18). While hands-on training alongside expert instructors is a useful and necessary component of echocardiographic education, our tool is attempting to build interpretive experience without requiring the space and time constraints of hands-on training. After using our online educational tool that allows users to learn on their own time and pace, trainees were able to improve their ability to correctly identify varying categories of LV systolic function, both at extreme and intermediary...
normal and abnormal cardiac function. Proficiency in echocardiographic assessment of ventricular function is developed by repetitive exposure and the creation of pattern recognition to appreciate subtle changes in the myocardial wall motion. Indeed, even experienced clinicians may benefit from an educational tool like ours. In our study, the average test score of experts was only 70% ± 10.61%, which provides further evidence that qualitative assessment of LV function is a difficult task even for experts. One study found only 50% agreement in qualitative assessment of LV function among experienced operators on a given image (20). Our tool can possibly help provide further training for experts to improve agreement.

Limitations

We note the following limitations in this study:

1. A potential confounding factor in evaluating the effectiveness of this educational tool is that the trainees, who were participating in a 3-week course that introduced echocardiography during the study period, may have accessed other echocardiography educational materials during the interval between the pre- and post-tests. However, the average time length between completion of the pre-test and completion of the post-test was 6 days, and the course taught image acquisition as part of a broader introduction to anesthesia with no specific training or formal curriculum on assessment of LV function. In addition, the trainees were non-clinical during the course, so they did not perform any bedside transthoracic echocardiography (TTE) during the study period. We therefore believe that these training modules were the primary modality of teaching assessment of varying levels of LV function for the interns during the time of the study.

2. Another potential confounding factor is that some interns had prior experience and/or training in TTE before participating in the study. One had completed a month-long course during medical school that focused on TTE and focused assessment with sonography in trauma (FAST). However, all other interns either had no formal training in TTE or were only introduced to basic TTE before participating in the course. We therefore believe that these training modules were the first formal modality of teaching assessment of LV function for most of the interns. To note, the intern who completed the month-long course in TTE and FAST exams scored higher on the post-test (70%) than the pre-test (45%). This result implies that despite having prior training in TTE, the intern’s ability to assess LV function still improved after completing our training modules.

3. This study focused on improving the intern’s ability to perform qualitative assessments of LV function, which is not as desirable as quantitative assessments. However, though it is not preferred, qualitative assessment is widely used in clinical practice with point-of-care echocardiography, so we believe teaching trainees this skill has practical value.

4. Our educational tool taught assessment of LV function using TTE clips, not transesophageal echocardiography (TEE) clips, which anesthesia residents use frequently in the intraoperative setting. While TEE is most useful in the intraoperative setting, TTE is becoming increasingly used in the perioperative setting, specifically in the preoperative setting, postoperative anesthesia care units, and intensive care units (21, 22).

5. In order to minimize any selection bias from the pre- and post-test design process, we administered the test to experts to compare intern performance to a standard, acceptable skill level for attending anesthesiologists specialized in cardiac anesthesia and/or critical care. Although we found that the interns scored within one S.D. of the experts’ average score or higher, further research should focus on validating this tool with a larger sample size of experts and with different groups of learners.

6. Although our sample size was enough to detect a difference in overall performance (our primary analysis), it is possible that we did not detect differences in performance by function level (our secondary analysis) because we were underpowered with our given sample size. Future studies should include larger sample sizes for further investigation on our secondary analysis.

7. It is not clear from this investigation as to how long lasting the improvement in identifying varying levels of LV function persists, whether or not it is necessary to have refresher training, or how often refresher training would be necessary. Further investigation involving comparison to expert performance and the persistence of training effects could be helpful in optimizing the educational tool. Expansion of this approach to other areas of cardiac assessment could also be useful.
Conclusion

There are many educational models available for ultrasound training, which take advantage of live workshops and web-based teaching. Most of these initiatives focus on a comprehensive educational program to develop a pre-determined level of competence. It can be a significant logistical challenge to ensure maximal trainee participation and minimize attrition. Our educational tool is unique in that it overcomes some of these logistical challenges. The content is independent of time and space constraints, and individual trainees can access the interactive information on their own time. High quality, anonymized echocardiography videos can be compressed and rapidly delivered world-wide to individual hand-held mobile devices without compromising patient privacy.

Left ventricular systolic function is often one of the key echocardiographic questions that need to be answered in the perioperative period and it can be one of the most difficult to master. We hope that the promising performance of this educational resource can be translated into more time- and cost-effective methods for improving diagnostic accuracy among learners.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

Acknowledgements

The authors would like to thank the Department of Anesthesia, Critical Care and Pain Medicine and the Center for Anesthesia Research Excellence (CARE) at Beth Israel Deaconess Medical Center for their support of the study. They would also like to thank the anesthesia residents who participated in the study.

References


5. Chu LF, Ngai LK, Young CA, Pearl RG, Macario A & Harrison TK. Preparing interns for anesthesiology residency training: development and assessment of the successful transition to anesthesiology residency training (START) e-learning curriculum. *Journal of Graduate Medical Education* 2013 **5** 125–129. (https://doi.org/10.4300/JGME-D-12-00121.1)


15. Bustam A, Noor Azhar M, Singh Veria R, Arumugam K & Loch A. Performance of emergency physicians in point-of-


Received in final form 3 December 2019
Accepted 17 February 2020
Accepted Manuscript published online 17 February 2020