Original Research

Considerations in Using Videoconferencing to Assess Multi-learner Classes of Layperson CPR: A Benchmark Study

Adam Gesicki[1], Ada Chan[1], and Sarah Witiw[1]

[1] Canadian Red Cross

As COVID-19 pushed many activities into the virtual space, many CPR training agencies were left to consider the role that videoconferencing may play in providing learners the opportunity to complete their learning. However, this challenged program managers to consider a pivotal question: Can facilitators of first aid effectively assess skills through videoconferencing? In this study, 45 instructors were exposed to mock videoconferences of layperson CPR learners, consisting of two classes each of 1-, 4-, 9-, and 12-learners in size. Overall, study participants were able to correctly identify the specific errors made by learners – or correctly make a declaration of no error – only 38.5% of the time. When the criterion was broadened to a pass/fail decision (i.e., correctly identifying that there is an error – regardless of the specific nature of the error), performance improved to 53.8%. Although this accuracy may not be acceptable for certain learning contexts, it should not be used as the sole reason of withholding CPR education in the virtual space across-the-board. An unstructured discussion held with study participants during the study provides insight into further considerations that should be accommodated for in future work aimed at improving instructor performance in the videoconferencing space.

The COVID-19 pandemic has spurred educators worldwide to consider new strategies for first aid education (as raised by Wilp, 2020). During this time, many program managers and educators were challenged with considering alternative ways of reaching out to learners, such as using videoconferencing platforms: a strategy that was quickly identified as being able to maintain a structure similar to a classroom setting, despite any restrictions such as physical distancing (Ramadan, 2020).

At first glance, some may consider videoconferencing as interchangeable with in-person experiences; particularly due to the prominence of many other activities transitioning online during COVID-19 (such as workplace meetings, academic conferences, or school-based educational experiences). However, some key differences between the in-person and videoconference-based experience – as it relates to the assessment of CPR – make it difficult to consider the two modes as exactly equivalent. These differences include: a lack in consistency in the visual backgrounds of learners, a lack of the ability for a facilitator to change their viewpoint, and the lack of directional audio cues. The question posed by the authors in this work was to illuminate the relevant considerations to be accounted for in the online space, and the degree to which the factors outlined above (or others) may play into the quality (accuracy) of the assessment.

There are two threads from the research literature that are worthy of consideration in light of this work; namely, that of instructor assessments of CPR in general, as well as the assessment of skills by way of video. Starting with the first theme, the current literature has already highlighted some of the challenges of relying solely on instructor assessments of CPR quality in the in-person mode. For example, Hansen, Bang, Stærk, Krogh, and Løfgren (2019) noted that instructors in their study assessed BLS learners with "high sensitivity but remarkably poor specificity" (p. 284). The quality of chest compression rate assessment was noted to be better than depth assessment; however, the authors theorized that this may be related to the potential use of time-keeping devices by the instructors (which was not recorded as a study variable). A similar study by Brennan, McGraw, and Brooks (2016) also highlighted disagreements between mannequin- and instructor-based assessment. Unlike the study by Hansen and team, Brennan and colleagues commented that rate was the "least accurate assessment" (p. 279). Despite variations in the specifics of instructor assessment, both studies highlighted a trend for false positives; in other words, instructors are more likely to pass a learner despite objective measures suggesting much poorer performance. It is worth noting that current work on this theme appears to be limited to the professional context, as opposed to the lay rescuer context. Although the challenges of unaided instructor assessment have been well-documented in the literature, including a systematic review conducted by ILCOR (Duff et al., 2021), this study was intentionally designed without the inclusion of feedback-enabled mannequins. This was a deliberate choice, with the intention of the study forming a baseline of facilitator performance, as well as respecting business considerations where the issuance of feedback-enabled mannequins to learners' homes may be considered unfeasible. It also mirrored the experience of instructors enrolled in the study, who (as of writing), were not required to use feedback-enabled devices in their layperson classrooms.

The current literature base has also started to explore the comparison of in-person versus videoconference-based assessment of resuscitation performance. For example, Weeks and Molsberry (2009) compared the evaluation of Paediatric Advanced Life Support (PALS) course participants by an instructor who was present in the room compared to an instructor who was watching the recording of the learners' performance. The study demonstrated the feasibility of evaluating participants in this manner, showcasing perfect agreement for pass/fail decisions amongst three evaluators for 26 out of 27 study participants. However, the nature of assessment in this study is different in nature compared to the work cited earlier in this paper. For example, a critical performance step in the Weeks and Molsberry study required participants to direct the preparation of an appropriate dose of epinephrine: an activity that provides the instructor with both verbal and physical cues to aid in their assessment. This is different from the motor evaluation that is required for the assessment of CPR (e.g., assessing the quality of compression depth). Expanding the literature search more broadly reveals how much video-based review has permeated health professional education, as it is considered one of many potential strategies to aid in the provision of feedback to learners (see INACSL Standards Committee, 2016 for an example of its use in nursing simulation; or the review conducted by Yanes and colleagues, 2015 for its application in broader medicine; or an example study such as the one performed by Rolston et al., 2020 for the use of video-based review in actual resuscitative scenarios). Once again, many of these studies were performed in professional contexts, with partial applicability in the layperson context. As well, these studies are limited in that they have focused on the single clinical care context. Even with multiple professionals working together in an advanced-level program, the evaluator's work is eased by only needing to consider one clinical scenario. In the context of layperson CPR classes, one instructor typically evaluates multiple learners at the same time, each performing within their own scenario (despite the 'story' that may be shared amongst the learners). This limitation was another driver supporting the need of further research.

Given these gaps in the literature, the study authors considered the following main question: What considerations do training agencies need to take into account before considering the integration of virtual assessment of CPR skills into their programming? More specifically, this study was intended as a benchmarking study; exploring the practicality of the assessment (from the facilitator's perspective), as well as the accuracy of facilitator assessments of performance, as it relates to multi-learner classrooms.

Methodology

Study Materials

Simulated Videoconference-Based Classes

To create classes of simulated learners, 12 study confederates were videorecorded performing CPR of varying quality. To ensure the recordings captured the intended performance, study confederates were either CPR instructors themselves or were supervised by two CPR instructors; all performance was validated using feedback-enabled mannequins; and, all recordings were cross-validated by the study authors one month after the recordings were taken. The videorecording exclusively focused on the provision of compressions and breaths (i.e., excluding any assessment sequence), and was bookended with 10 seconds of inaction to allow study participants to become accustomed to the variety of settings the learners were in. The recordings featured a variety of backgrounds that would resemble CPR practice in the home setting.

The confederate videos were then assembled to form a sequence of simulated videoconferences. The classes were administered in ascending class sizes of 1, 4, 9, and 12 confederates. Each class size consisted of two independent trials, totaling eight simulated classes for study participants to assess. Although confederates were unique within each class, confederates reappeared between classes (study participants were specifically instructed to disregard this during study orientation). For clarity, each video with an error was only used once throughout the entire study; however, some error-free learner performances were repeated throughout the study (further discussed in the Results section, below). The final errors that were included in each class are summarized in Table 1, and a sample screenshot of a 12-learner class is shown in Figure 1.

Survey and Data Collection File

A survey was also designed to better understand the participants' experience, both in terms of how often they taught CPR programs (before various COVID-19 restrictions took effect), and the number of years that they have been instructing CPR. Furthermore, study participants were provided questions regarding three variables that were hypothesized to have an impact on study performance; namely, the size of their monitor, their default body position while assessing in the classroom, and how they traditionally move while assessing in the classroom.

Accompanying the survey was the data collection file. The file provided a 'square' for each confederate that the study participant would view during the recording, consisting of checkboxes of common errors as well as a free-text field (see Figure 2). The file was laid out in the same manner as learners would appear on their screen.

Study Administration

Study participants were eligible to take part in the study as long as they were certified to teach CPR Level C (layperson CPR) for the Canadian Red Cross, had high-speed internet, and were able to use a monitor larger than 12". Each participant was provided with a consent form in advance of

Table 1
A listing of the confederates' performances in each simulated videoconference/class

Class #	Number of Learners	Learner Performances
1	1	• 1 confederate performed error-free CPR
2	1	• 1 confederate made no attempt at head-tilt chin-lift
3	4	• 3 confederates performed error-free CPR • 1 confederate slowed down to a compression rate of 90bpm in rounds two, three, and four (of five rounds)
4	4	• 2 confederates performed error-free CPR • 1 confederate rocked the head back and forth, instead of administering a breath • 1 confederate slowed down to a compression rate of 95bpm after the first round
5	9	• 7 confederates performed error-free CPR • 1 confederate had excessively long hands-off time • 1 confederate sped up to a compression rate of 130bpm in rounds two, three, and four (of six rounds)
6	9	• 6 confederates performed error-free CPR • 1 confederate's compressions were too shallow • 1 confederate did not compress in the centre of the chest (positioned superiorly from ideal position) • 1 confederate appeared to have a head-tilt chin-lift, but the airway was not open enough for breaths to go through
7	12	• 9 confederates performed error-free CPR • 1 confederate provided breaths that were too full • 1 confederate did not compress in the centre of the chest (positioned laterally from midline) • 1 confederate provided compressions that were too deep
8	12	• 8 confederates performed error-free CPR • 1 learner did not compress in the centre of the chest (positioned inferiorly from ideal position) • 1 confederate's breaths did not always go in • 1 confederate's compressions were too shallow • 1 confederate was not wearing gloves

registering for this study. Participation was on a voluntary basis. The study was administered by videoconferencing platform Zoom and lasted an hour. In the first fifteen minutes, a study author (AG) re-introduced the purpose of the study and answered questions about the consent form. Participants were then introduced to the data collection tool and were directed to use it to note errors made by confederates. An error was defined as "something that you would give corrective feedback about to the learner [confederate]". Once the participants were ready to proceed, the researcher began to cast the study video. The study video was paused in-between classes to provide enough time for study participants to note their feedback on the data collection form, taking approximately 30 minutes to complete. At the completion of the video, the researcher invited study participants for an unstructured discussion.

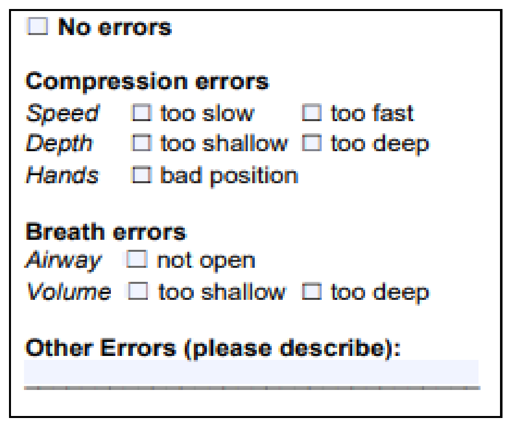
Data Processing and Scoring

Notes of the unstructured conversation were taken during the conversation by one study author, while another study author listened to validate notes taken. Thematic analysis was done

Figure 1 *A screenshot of Class #8 (a 12-confederate classroom)*

by consolidating the notes taken from the unstructured discussion as well as any e-mailed feedback provided by study participants.

In addition to the thematic notes, two scoring strategies were used to interpret the data collection files: *assessment for learning* (AfL) and *assessment of learning* (AoL). The purpose of the AfL strategy was to identify whether the study participant was able to correctly identify the specific error being made by the confederate in the video. As such, a score of 1 was awarded to a participant if they were able to correctly specify the exact error that were made by a particular confederate; or, marked "no error" where there indeed was no error. A score of 0 was awarded otherwise. The AoL measure awarded the study participant a score of 1 if the participant identified an error where there indeed was an error in the confederate's recording (regardless of the specific error); or, if "no error" was checked by the study participant where there was indeed no error in the confederate's performance. Under this scoring strategy, a score of 0 was awarded when the study participant checked "no error" for a confederate who was making an error or vice-versa. These two scoring strategies contrast the two roles an instructor plays during an educational intervention: the role of guide and mentor, who needs to provide adjustments to learners with an aim to improving performance (i.e., AfL), as well as the role of evaluator and reporter, who needs to make a pass/fail decision on the learner's final performance (i.e., AoL). The terminology used here is informed by the educator roles and approaches presented by Earl (2003), with acknowledgement that the role of the educator can be assessed at a much deeper level than was adopted for this initial line of research. The specific analytical techniques used to analyze



☐ **No errors**

Compression errors
Speed ☐ too slow ☐ too fast
Depth ☐ too shallow ☐ too deep
Hands ☐ bad position

Breath errors
Airway ☐ not open
Volume ☐ too shallow ☐ too deep

Other Errors (please describe):

Figure 2 A sample "square" that an instructor would use to record learner performance participant scores vary by research question and are addressed further in the paper.

Results

Study Participation and Survey Results

Overall, 50 participants consented to participation and attended one of the ten study sessions that were offered. Of the 50 participants, 45 participants submitted useable confederate ratings (the difference of five were due to technical issues with study administration or data transmission). Three of the 45 instructors submitted their ratings without an accompanying survey file. Survey data revealed that study participant experience was predominantly skewed towards individuals with more years of instructor qualifications; however, experience by frequency of courses taught was more evenly distributed.

Themes Identified in the Unstructured Discussion

The themes from the unstructured discussion reached saturation by the fifth (of ten) study administrations. The strongest theme that was identified was the role that color, and contrast played in the participants' (perceived) ability to assess in the study. Participants expressed that glare was a significant issue in particular recordings, especially when the mannequin skin color was lighter, and an overhead light shone directly on the mannequin. Three participants specifically noted that their assessments were helped when there was a contrast between the learner, the mannequin, and the flooring/background. Further in the theme of contrasts, participants were particularly drawn to colored gloves used in the study, with some participants only learning during the discussion that some confederates wore translucent gloves (mistaking it for the confederate not wearing gloves at all). Distracting elements, such as a cat laying next to the mannequin – even though they did not interfere with watching the skill – were identified as significant barriers to quality assessment.

The next most emergent theme was a discussion on ideal class size. Participants imagined feeling most comfortable by assessing six to eight learners at once on their screen; often describing the jump between four and nine confederates as being too great for comfort.

The majority of participants also felt a head-on camera angle would have been easier for assessment, with two participants suggesting multiple camera angles per confederate would allow for optimal assessment. Overall, there was much appreciation in study participants in the key differences between online assessment and in-person assessment. Participants felt that the power to move and change their view in the in-person setting added significant value over the static angle presented in the study.

Scoring Results

The overall median AfL score across instructors was 38.5% (IQR 28.4% – 46.2%), and the median AoL score was 53.8% (IQR 48.1% – 61.5%). Given that more confederates had error-free performance in the study than those containing errors, a confusion matrix was created to better illustrate participant behavior (Table 2).

The confusion table reveals that study participants accurately identify correct performance (true positive rate/sensitivity) in 50.7% of cases, and a correctly identify incorrect performance (true negative rate/specificity) in 65.4% of cases.

Table 2

Confusion matrix for Assessment of Learning (AoL) ratings between instructors (study participants) and learners (study confederates).

	Rated Incorrect	Rated Correct
Actually Incorrect Performance	471 [TN]	249 [FP]
Actually Correct Performance	798 [FN]	822 [TP]

Next, participant-level scores were broken down by class size and are displayed visually in Figure 3. The results show that a participant's accuracy for identifying specific errors being made by confederates appeared to increase as class size increased (medians of 0%, 25%, 38.9%, and 45.8% for class sizes of 1, 4, 9, and 12 learners respectively).

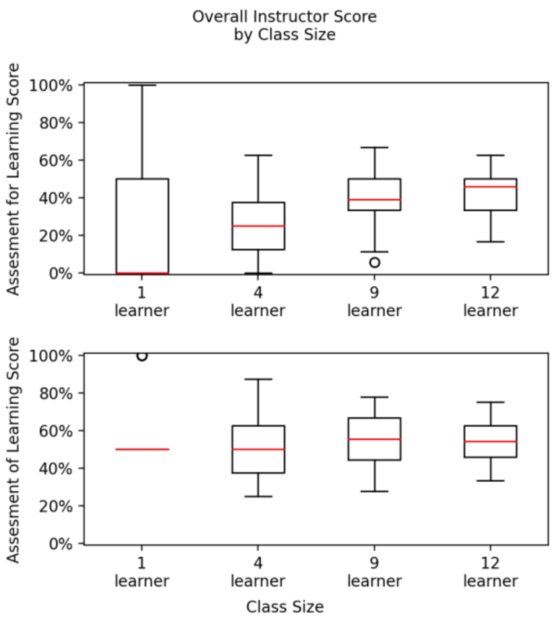


Figure 3 Instructor performance across class sizes (each class size had two classes' worth of learners).

The survey results were then used to contextualize some of the results. First, participants' monitor size was median split into two groups: the "small" monitor size group for monitors up to and including 16 inches across ($n = 24$), and the "large" monitor group consisting of the remainder ($n = 20$). This resulted in a AoL medians of 55.8% and 48.1% respectively (Mann Whitney $U = 125.5$, $p < 0.05$). Classroom assessment practices also had a statistically significant result: 28 study participants reported preferring to stand while assessing learners perform CPR (median AoL score 54.8%); versus 14, who default to kneeling (median 50.0%; Mann Whitney $U = 118.0$, $p < 0.05$). Participant experience, as defined by groupings of the number of courses taught per year, did not play a significant factor in assessment results (Kruskal-Wallis = 2.05, $p > 0.05$); neither did curriculum revision that the study participant first certified under (Kruskal-Wallis = 2.42, $p > 0.05$).

As mentioned in the Study Materials section, some of the error-free confederate performances were repeated multiple times throughout the eight classes; in particular: two confederates appeared in two different classes, five confederates appeared in three different classes, and four confederates appeared in four different classes. These 11 confederates shown in multiple classes permitted for an analysis of intra-rater consistency. For a study participant to be considered consistent for a particular confederate, their AoL scores for the confederate were expected to be identical across all the virtual classes (regardless of the confederate's actual performance in the study video). Under this definition, two participants rated 2/11 consistently, three participants rated 3/11 consistently, fourteen participants rated 4/11 consistently, eleven participants rated 5/11 consistently, eight rated 6/11 consistently, six rated 7/11 consistently, and only one participant rated 8/11 consistently.

In keeping with the work already established in the literature (Hansen et al., 2019; Brennan et al., 2016), participant sensitivity to compression speed was considered next. Speeds demonstrated by the confederates were intentionally varied throughout the study: the number of confederates with each speed, as well as participant judgements of each speed, is indicated in Table 3. As the Table shows, most participants were sensitive to slow speeds only once it dropped to 90 compressions per minute (cpm); and for the most part, did not flag 130 cpm as too fast. This is despite the current CPR curriculum expecting learners to perform compressions within the 100-120 cpm range (Canadian Red Cross, 2017).

Table 3
Number of learners demonstrating each compression speed (as rows; grouped by dominant speed), and the proportion of instructor ratings for each error (as columns). Majority groups have been bolded.

	number of learners shown at speed	too slow	no error	too fast
90 cpm	1	**64%**	33%	2%
95 cpm	1	25%	**71%**	4%
105 cpm	10	5%	**94%**	1%
110 cpm	23	11%	**85%**	3%
115 cpm	13	5%	**87%**	7%
120 cpm	3	3%	**87%**	10%
130 cpm	1	2%	**89%**	9%

In addition to speed, previous literature has commented on instructor perception of depth, which deserves special attention in this study's results. Notably, study participants frequently recorded depth as an error made by confederates, despite it not being an error being displayed by the confederates. The "too shallow" error was marked by study participant 542 times across the 2340 participant-by-confederate combinations, even though the error was only ever shown in two confederates. 495 of these 542 ratings were provided to confederates with appropriate compression depth, leaving only 47 ratings as true "too shallow" ratings. No "too shallow" ratings were provided to the confederate who was compressing too deep.

Discussion

Overall, participant performance in this study was certainly below a threshold that could be relied on in situations requiring skill validation with high certainty, despite perceived participant comfort with six-to-eight learners [confederates]. This study especially highlights the challenges instructors face in assessing depth of compressions, an effect that is likely exacerbated by the forced perspective of a single camera angle. However, unlike previous work (Hansen et al., 2019; Brennan et al., 2016), instructors in this study were more likely to identify poor performance where the learner's performance was not poor. This non-conforming result cannot be easily explained with the data collected to-date in this study. Given the benchmarking nature of this work – namely being the first highlighting virtual assessment in multi-learner CPR classrooms – it would certainly benefit from further study.

The counter-intuitive results related to class size (i.e., more learners were associated with more accurate assessment) and monitor size (i.e., larger monitors were associated with less accurate assessment) were also of particular interest. Some of this may be explained by the effect expressed by a few participants in the unstructured discussion: classrooms with fewer learners may make instructors more attentive in spotting errors than larger classes. Naturally, a consistent evaluation benchmark should be expected across all participants, regardless of class size. This underscores the importance of considering class size in any future studies assessing the evaluative practices of facilitators. The counter-intuitive effect related to monitor size may be explained by the same reasoning, even though it was never explicitly stated: since a smaller monitor would mean each individual participant was rendered in a smaller area, instructors may not have been as perceptive of errors as they would be on a larger screen. The extent to which assessor training programs (such as the one outlined by Thorne and colleagues, 2013) may combat these behaviors may be worth considering. Further research may certainly open the consideration of optimal, or maximum, learner-to-facilitator ratios to enable accurate skill evaluation.

Study participants were quick to discuss the role that alternative camera angles could play in assessment – a theme that is in-line with previous work (Jones et al., 2015). Many pointed out that they would have preferred a head-on camera angle instead of the top-down angle, as it would better showcase compression depth and breath depth.

Although this camera angle may solve the issue of monitoring compression depth and breath depth, instructors would then lose the landmarks required to assess hand position on the chest fully, or to be able to visualize issues with mask seal. Given that study participants recognized the value of being able to move around in the in-person setting, this discussion theme may be reflective of instructors recognizing what they are missing to complete the assessment – without explicitly recognizing the benefits that the top-down camera angle contributed to their assessment.

The survey question on the study participant's default body position was intended to confirm the authors' hypothesis that the top-down angle camera angle used in this study would most resemble instructing practice, and therefore be more familiar to participants. Although the survey results suggested that most participants prefer to stand when evaluating learners, this is quite different from the head-on angle suggested by participants in the study's unstructured discussion. This key difference suggests that there may be deeper differences in the assessment of performance in the in-person and online modes; differences that may need to be addressed with targeted facilitator training to enable accurate assessment of skills in both spheres.

A solution that may come easily to mind is the potential of feedback-enabled ('smart') mannequins in addressing some of the gaps identified in the study. As mentioned earlier, this may be cost-prohibitive in specific markets. A post-hoc analysis was considered on this dataset to reflect the potential effect a feedback-enabled mannequin could have on instructor performance. However, this was not performed as it would not be realistically informative. In the positive case, a feedback-enabled mannequin might lessen an instructors' cognitive load, allowing them to observe other errors more clearly. In the negative case, allowing instructors to limit their attention on fewer aspects of performance would be counter-productive with groups of overly stringent instructors. The question of how best to integrate feedback devices in instructor teaching remains open to ongoing research (Duff et al., 2021), and as such, the assessment of the interface between instructor and feedback-enabled mannequin is best left to studies designed with specific intent to address that line of research.

Future work could also further investigate the effect that marking grids may have on instructor performance. Since the marking grid presented to study participants in this study (as was pictured in Figure 2) contained explicit mention of the most frequently demonstrated errors made by learners in CPR, participants may have been unconsciously cued to be more attentive to indicators of *error*, as opposed to indicators of *successful performance*. This hypothesis may partially explain the overly stringent performance of instructors in this study, when compared to trends previously documented in the literature. This effect does not seem to have been buffered by the verbal instructions provided to study participants, asking them to focus on corrective feedback they would provide learners in the classroom. As such, further experimental work could also explore the extent to which cuing provided to an evaluator may influence their scoring of CPR performance.

Lastly, it is important to acknowledge there are other challenges that are introduced with the use of videoconferencing for learning. These include the impact of a learner's socioeconomic status, access to broadband internet, and their technical literacy surrounding the use of videoconferencing. These aspects are important to consider from a training agency and learner perspective before the implementation of videoconferencing as a learning and skill assessment modality (see Wilp, 2020).

Conclusion

In sum, the research on instructor performance in assessing CPR using videoconferencing is still in its infancy. Despite the somewhat expected nature of these findings – given previous literature of in-person instructor assessment performance – this study provided a baseline for identifying several other factors that are relevant to consider in future approaches to online skill assessment. Although the performance of instructors in the virtual space is less than ideal, it is naïve to discount this learning modality as a future approach to learning; particularly given its explosive uptake during the COVID-19 pandemic, and the benefit it provides for increased learner outreach. Naturally, program developers must be mindful in their decision to integrate videoconferencing into their learning programs, balancing the potential benefit of increased learner outreach, with the risks that it creates to program quality. More explicitly: remote communities may benefit from a virtually connected facilitator, especially if it meant they would be able to receive education where they would not otherwise have. However, this modality may be less applicable in contexts where technical mastery is expected or required. The understanding that drives this decision will, of course, shift over time as this line of inquiry continues to evolve.

Acknowledgements & Funding

The study authors deeply acknowledge that this study, as well as the work of the Canadian Red Cross, occurs on traditional, ancestral, and unceded territories of many Indigenous Nations. The virtual nature of this study drew together individuals from territories spanning coast, to coast, to coast.

The study authors would like to thank those who participated in this study, either confederates or study participants. In addition, the thoughtful reflections of JM and KW for the first draft of this article are sincerely appreciated.

No additional funding was used for this project.

Conflict of Interests

The study authors have no conflict of interests to report.

Corresponding Author

Adam Gesicki, adam.gesicki@redcross.ca

References

Brennan, E. E., McGraw, R. C., & Brooks, S. C. (2016). Accuracy of instructor assessment of chest compression quality during simulated resuscitation. *Canadian Journal of Emergency Medicine*, *18*(4), 276-282. https://doi.org/10.1017/cem.2015.104

Canadian Red Cross. (2017). *First Aid & CPR.* Retrieved from http://www.redcross.ca/facpr

Canadian Red Cross. (2019). *First Aid Program Standards: June 2019.* Retrieved from myrc.redcross.ca (requires instructor log-in credentials).

Duff, J. P., Shammet, S., Damjanovic, D., Bhanji, F., Bigham, B.L., Bray, J. E., Breckwoldt, J., Cheng, A., Glerup Lauridsen, K. G., Gilfoyle, E., Hsieh, M.J., Iwami, T., Lockey, A.S., Ma, M., Monsieurs, K.G., Okamoto, D., Pellegrino, J.L., Yeung, J., Finn, J., Greif, R. - on behalf of the International Liaison Committee on Resuscitation Education, Implementation and Teams Task Force. (2021). *CPR Feedback Devices in Training (EIT #648): Systematic Review*. http://ilcor.org

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to enhance learning.* Thousand Oaks, CA: Corwin Press.

Hansen, C., Bang, C., Stærk, M., Krogh, K., & Løfgren, B. (2019). Certified Basic Life Support instructors identify improper cardiopulmonary resuscitation skills poorly: Instructor assessments versus resuscitation manikin data. *Simulation in Healthcare*, *14*(5), p. 281-286. https://doi.org/10.1097/sih.0000000000000386

INACSL Standards Committee. (2016). INACSL standards of best practice: Simulation[SM] Debriefing. *Clinical Simulation in Nursing, 12*(S), S21-S25. https://doi.org/10.1016/j.ecns.2016.09.008

Jones, A., Lin, Y., Nettel-Aguirre, A., Gilfoyle, E., & Cheng, A. (2015). Visual assessment of CPR quality during pediatric cardiac arrest: Does point of view matter? *Resuscitation, 90*, 50-55. https://doi.org/10.1016/j.resuscitation.2015.01.036

Ramadan, T. (2020). Adapting the First Aid education and response to the COVID-19 pandemic crisis. *International Journal of First Aid Education, 3*(1), 9. http://dx.doi.org/10.21038/ijfa.2020.0103

Rolston, D. M., Li, T., Owens, C., Haddad, G., Palmieri, T. J., Blinder, V., …, Becker, L. B. (2020). Mechanical, team-focused, video-reviewed cardiopulmonary resuscitation improves return of spontaneous circulation after emergency department implementation. *Journal of the American Heart Association*, *9*(6). https://doi.org/10.1161/JAHA.119.014420

Thorne, C. J., Jones, C. M., Harvey, P., Hulme, J., & Owen, A. (2013). An analysis of the introduction and efficacy of a novel training programme for ERC basic life support assessors. *Resuscitation*, *84*(4), 526-529. https://doi.org/10.1016/j.resuscitation.2012.09.030

Weeks, D. L., & Molsberry, D. M. (2009). Feasibility and reliability of remote assessment of PALS psychomotor skills via interactive videoconferencing. *Resuscitation*, *80*(3), 354-358. https://doi.org/10.1016/j.resuscitation.2008.11.025

Wilp, T. (2020). Learning, teaching and first aid in times of COVID-19*. International Journal of First Aid Education*, 3(1), 5. http://dx.doi.org/10.21038/ijfa.2020.0102

Yanes, A. F., McElroy, L. M., Abecassis, Z. A., Holl, J., Woods, D. & Ladner, D. P. (2015). Observation for assessment of clinician performance: A narrative review. *BMJ Quality & Safety*, *25*, *46-55.* http://dx.doi.org/10.1136/bmjqs-2015-004171