

**Peer Review File**  
**Article information: <https://dx.doi.org/10.21037/jphe-23-89>**

**Reviewer A**

Comment: (1) Adding more clearly that the null hypothesis is the equality of populations and that the alternative is the difference in populations would help the reader.

Reply: We fully agree.

Changes in the text: 144

Comment: (2) To help the reader, “P-value = 0.047” should be written instead of “P = 0.047” in all similar its occurrences.

Reply: We fully agree.

Changes in the text: 144, 156, 159, 162, 170

**Reviewer B**

Comment: The manuscript refers to the experimental manipulation in different places as “time pressure” or “verbal pressure.” I recommend using “verbal pressure” uniformly since a time limit was not imposed.

Reply: We fully agree.

Changes in the text: **Highlight Box**, 26, 28, 55, 85

Comment: Page 4 Line 77. Should “stuff” be “staff?”

Reply: We fully agree. Excuse us for this typo.

Changes in the text: 81

Comment: Page 5 Line 91. What training did the subjects have in the tasks they were asked to perform. Did they have experience with each task in patients and did the groups differ?

Reply: The SPEED-BOMB checklist were introduced them previously but they did not exercised such a scenario previously. No differences between groups (all subjects are from the same students group). They are of course trained in good quality of CPR.

Comment: Line 100. Did the leader follow a script to urge faster kit preparation? What was done to assure that the degree of urgency was constant from one subject to the next?

Reply: During all scenarios the leader was the same person and was using the same words/phrases prepared previously.

Changes in the text: 104 we add ‘using previously prepared words and phrases’

Page 6 Line 145. Text reads, “the study group accomplished the task with a lower sum of study points, which means they were more effective.” It’s not clear what “more effective” means in this context and whether the data

support the statement. I would say that the two groups were indistinguishable at CPR depth and rate, but the study group used less time in preparing the intubation kit. I'm unsure whether the point scale is a valid measure of performance.

Reply: We suggest following changes: In general, the study group accomplished the task with a lower sum of points (despite similar chest compression quality and accuracy of intubation kit preparation so the majority of differences between groups is connected with time).

Changes: as above 150-153.

Page 9, Line 218, should read "heart rate" instead of "blood pressure."

Reply: We agree.

Changes: 225.

Comment: Page 9, line 228. The investigators reported no data about subjects' physical condition and cannot draw any conclusions about the impact of physical condition on the rate of chest compressions.

Reply: Polish paramedic students in general take part in some physical activities (during their education) and they are in general fit and healthy. There were no features suggesting any alterations over this fact in both groups' members.

#### Major Comments

Comment: Lines 91-105. The scenario is somewhat artificial from my perspective. Equipment and supplies for emergency intubations in a hospital are generally prepared and assembled proactively. Early intubation is not a part of current ACLS guidelines. Also, the two tasks are independent of each other and not closely related in training and skills needed. What was the rationale for evaluating intubation kit preparation combined with performance of CPR.

Reply: The intubation equipment location differs widely between departments and it is not clear in all Polish EDs if it should be assembled proactively. The scenario does not need to be strictly ACLS related one (there is no cardiac arrest at the beginning). Even if endotracheal intubation is not crucial according to ACLS guidelines, there are some cases when it could be a better idea than supraglottic devices (especially before cardiac arrest). Above details seem not to influence the study which is about changes in personal behaviour during the scenario (if verbal stress is added or not). Combining both tasks in one scenario could give us information about influence of verbal stressing on actual but also further activities (even if the stressing stops like during CPR but may affect its quality).

Comment: The scoring system for evaluating intubation kit preparation is arbitrary. How did the investigators determine the time penalties for deficiencies in preparing the intubation kit and quality of CPR? Has the scoring system been validated as a measure or predictor of some desirable outcome? Why didn't the investigators separate time and accuracy of kit preparation as independent assessments. What is the justification for combining elements from kit preparation and CPR in one score? These are independent, unrelated activities.

Reply: We appreciate your detailed analysis of this issue. We would like to focus your attention on Table 3 which contain time and preparation accuracy separately as well as CPR quality. You are fully right that the scoring

system is not validated till this time because it is authorial and was not used till this time. Our study is the pilot one and we are in the middle of further researches in which we are planning to validate and widen the scoring systems. Till this time, the scoring system is in fact arbitrary but it shows statistically significant differences between groups (even if groups are small) – the combination of all parameters shows us truly their better outcome (faster preparation without higher number of missing elements but also better chest compression). Our aim was to combine all parameters together (time, accuracy and CRP quality) as a 1 parameter for evaluating all the students' activities but all elements are of course shown separately. All of them are crucial in emergency situation – it is not only important to act fast but also prepare all the necessary equipment. High quality of CPR is vital. Moreover, it is not only important to resuscitate good and prepare all the equipment if it last a lot of time.

Comment: Table 3. Self-reported stress didn't differ between the groups. There is no evidence that experimental manipulation with verbal urging had the desired effect in raising stress. Figure 3 shows a wide range in self-reported stress, but one would surmise that this was a reflection of variation in the innate stress subjects experienced during the study. Different symbols should be used for experimental subjects and control subjects in Figure 3 so readers can see the impact of the verbal urging.

Reply: We fully agree.

Changes: 'Self reported stress' instead of 'stress' in Table 3. and Fig. 3. Different symbols were applied in Fig. 3. + proper changes in Figure description (167).

Comment: I agree with the investigators that the small number of subjects is a significant limitation in this study. They should conduct a power analysis to estimate how many subjects they would need for statistical significance given the effect sizes they observed.

Reply: Thank You for this valuable comment but please kindly notice that this paper is a pilot study before further publications. We are conducting researches with greater amount of participants and with different types of tasks (as a further research steps) so we do not see such a need at this point.