Reviewer's report

Title: Basic life support is effectively taught in groups of three, five and eight students. A prospective, randomized and double-blind study.

Version: 2

Date: 18 January 2014

Reviewer: peter iserbyt

Reviewer's report:

MS: 'Basic life support is effectively taught in groups of three, five and eight students. A prospective, randomized and double-blind study.'

Moritz Mahling, Alexander Münch, Sebastian Schenk, Stephan Volkert, Andreas Rein, Uwe Teichner, Pascal Piontek, Leopold Haffner, Daniel Heine, Andreas Manger, Jörg Reutershan, Peter Rosenberger, Anne Werner, Stephan Zipfel and Nora Celebi

BMC Medical Education
Research article
Reviewer: Peter Iserbyt

This study presents an interesting and practical research question within medical education. Determining appropriate group sizes is important to maximize learning, and is very much worth investigating. It has also important financial and organizational implications.

To me however, at this point, I feel the authors have not provided enough detail to make a good assessment of their work. In particular, I find the methods section too weak for this journal. There is simply not enough information provided. In general, the manuscript is very well written.

Please find my comments and suggestions below. I hope they will be useful for the authors and will serve the quality of this work.

Major Compulsory Revisions

Abstract:

1. In the first paragraph it is stated that both students and teacher prefer teaching in small groups. However, this statement is not made in the intro nor discussion (and thus not referenced).

2. I think it should be more stressed that the target population was medical students. As the authors state in the discussion: it is a limitation to the generalizability of findings that the medical students were in their 4th year, and consequently had prior knowledge (had high pretest scores). Therefore, the conclusions of the manuscript might not be valid for laypeople. I suggest that ‘medical students’ is added in the conclusion (last sentence) to stress this. The authors might also want to consider adding this in the title.
Methods:
1. I applaud the authors for permuting teachers to reduce teacher bias. However: how many students were taught simultaneously? Was there 1 teacher working with one group of 3, 5 or 8 at the time? How many groups did the teacher have to supervise? Were the groups all in one room? How many groups did 1 teacher had to teach? The authors should provide more detail here.
2. Was there some kind of instrument used to assess treatment validity: did the teacher do what was expected of him/her during the BLS class?
3. Was interrater reliability calculated based on data from all students or on a smaller sample?
4. Can the authors provide a rationale for their choice to compose groups of 3, 5 and 8? Why not 2, 4 and 6?
5. In the ‘Outcomes’ paragraph: if observers had a significant disagreement, they engaged in a discussion to find a consensus. However, why then is the interrater reliability not 100% (see results, second paragraph).
6. I think the authors should provide more detail about the role of the tutor: time management? Providing feedback? What was he or she asked or allowed to do and did he/she did that (validity)?
7. What was the basis to be able to detect a 4-point OSCE score difference between groups (as the basis for sample size calculation)?
8. What was the time interval between pretest, intervention, and posttest?
9. Was reliability on questions, conversations, hands-on time, and tutor interventions also assessed?

Results
1. I notice that no confidence intervals were provided.

Discussion
1. I encourage the authors to put more emphasis on their target group as a serious limitation to their work.

Minor Essential Revisions

Background
1. It might be useful to define what you mean by ‘simulation’ training, or its difference with BLS training (used interchangeably throughout the manuscript).

Methods
1. Please delete ‘one-way analysis of variance’ after ‘Kruskal-Wallis’
2. Statistical analysis, 2nd paragraph: not sure what is meant by ‘irr’ here. Interrater reliability?
3. Is a question/answer dialog strictly between tutor and trainee and a conversation only between trainees?
4. Although it is a good thing to switch tutors over different groups, there might
still exist a tutor-effect. Did the authors assess a possible tutor effect by investigating variance in test scores due to the teacher (e.g., by computing a 'null model')?

5. During the training in groups: what were students doing who were not practicing (hands-on)? Were they instructed to observe, were they just waiting?

6. Training time was 20 minutes introduction + 7 minutes practical training per student. Does that mean that class time in groups of 8 was 76 minutes?

7. Who read the instructions to the students for BLS assessment?

8. We first read about self-assessment in the methods. The authors did not provide a rationale or literature about this in the intro.

Results

1. Why didn’t you perform a questionnaire instead of a baseline test to assess prior knowledge? In addition, isn’t it possible that this baseline test activated inert knowledge? Or produced any type of learning effect?

2. I am not really convinced of the added value of conducting a pre self-assessment questionnaire to students.

3. As an educator, I am interested whether student scores within a group were correlated. Did you take look at this?

4. Is there other published research applying the 42 points pass rate? Or perhaps this is incorporated in the Brennan checklist itself?

5. I would suggest the authors to put IRR data in a first paragraph of the results, together with demographic data (table A) and the excluded datasets.

6. In table 1 we read ‘months since last resuscitation training’. This was not mentioned in the text, I believe.

Discussion

1. Do the authors consider ‘asking questions’ as a positive or negative thing? I assume it could be both and maybe the authors could provide a more balanced discussion about this.

2. Please use BLS abbreviated (last sentence 2nd paragraph) throughout the manuscript after it was written in full the first time you mention it.

Conclusion

1. The authors use terminology that in my opinion does not match the manuscript results. For example: ‘students benefitted from a longer practice time...’. It would say they did not benefit from that, since they achieved no higher learning. Also the authors suggest ‘..., but smaller groups might have additional learning benefits.’ I feel this statement has no place in the conclusion and should be addressed in the discussion (also: which additional benefits are do you mean)?

Discretionary Revisions

Figures
1. Not sure what happened with Figure 1 but its size is problematic (possibly due to upload/download).

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

'I declare that I have no competing interests' below.