Author's response to reviews

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

Authors:

Moritz Mahling (moritz@mahling.eu)
Alexander Münch (me@alexandermuench.eu)
Sebastian Schenk (s.schenk@student.uni-tuebingen.de)
Stephan Volkert (stephan.volkert@student.uni-tuebingen.de)
Andreas Rein (AndreasRein@t-online.de)
Uwe Teichner (uwe.teichner@med.uni-tuebingen.de)
Pascal Piontek (pascalpiontek@web.de)
Leopold Haffner (leohaffner@gmail.com)
Daniel Heine (daniel.heine@med.uni-tuebingen.de)
Andreas Manger (andreas.manger@med.uni-tuebingen.de)
Jörg Reutershan (joerg.reutershan@med.uni-tuebingen.de)
Peter Rosenberger (peter.rosenberger@med.uni-tuebingen.de)
Anne Werner (anne.werner@med.uni-tuebingen.de)
Stephan Zipfel (stephan.zipfel@med.uni-tuebingen.de)
Nora Celebi (nora_celebi@vertexbreeder.com)

Version: 3
Date: 14 April 2014

Author's response to reviews:

MS: 8952111941063830

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

Dear Dr. Ulep, Dear Cherrylyn Raytos,

we would like to thank you and the reviewers for the helpful comments. Please see the revised manuscript attached. Changes and relevant parts for our answers to the referees comments are highlighted in the manuscript.

Referee 1

Referee 1 comment 1: 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.

Authors’ reply: Thank you very much for the detailed and well-structured review of our study. We are happy to provide further methodological information, see below.

Major Compulsory Revisions

Referee 1 comment 2: 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).

Authors’ reply: There is some evidence that supports the preference of medical students for small group teaching, but interestingly not for teachers. We amended the abstract (page 3, line 3) and included the references in the background section (page 5, line 15-16).

---

Referee 1 comment 3: Abstract: 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.

Authors’ reply: We definitely agree, the influence on group size on the resuscitation abilities of lay people should be investigated separately. We therefore changed the last sentence of the abstract as well as the title and stress this point in the discussion. (page 4, line 1, and title)

---

Referee 1 comment 4: 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.

Authors’ reply: Teacher bias is also an important confounder in our opinion. We therefore assigned one tutor to the whole group of three, five or eight students. Each group had its own tutor and room, and the tutor did only supervise one group. The manuscript was modified to provide more details here (page 7, lines 15-16).

---

Referee 1 comment 5: Methods 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?

Authors’ reply: We strongly emphasized standardization prior to every teaching
session, and each tutor was supplied with a detailed, written teaching plan that
guided him through the course. We videotaped all the teaching sessions, there
was a fairly strict adherence to the teaching plan. We added this information in
the methods section (page 7, line 10-12). We further provide the teaching play as
additional file 2 (translated from German, illustrations have been removed due to
copyright limitations).

---

Referee 1 comment 6: Methods 3. Was interrater reliability calculated based on
data from all students or on a smaller sample?
Authors’ reply: Yes, calculations were performed based on the dataset of each
individual student.

---

Referee 1 comment 7: Methods 4: Can the authors provide a rationale for their
choice to compose groups of 3, 5 and 8? Why not 2, 4 and 6?
Authors’ reply: We hypothesized that students learned from either hands-on
experience or observation. Thus we needed three students as the smallest group
since during resuscitation, two students perform chest compression and
ventilation, the third student can observe. With only two students, there is no
opportunity for observation. For the group of three students, the ratio between
hands-on and observation is 2:1, for the group of eight students, the ratio is 1:3.
Five was the third group in between with a more balanced ratio of hands-on and
observation.

---

Referee 1 comment 8: Methods 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).
Authors’ reply: The consensus process was only sought when disagreement was
significant, i.e. below .50. All other ratings were accepted, therefore final
Interclass correlation did not reach 100%.

---

Referee 1 comment 9: Methods 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)?
Authors’ reply: The tutors received a detailed teaching plan. We added one as
supplemental material.

---

Referee 1 comment 10: Methods 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)?
Authors’ reply: We did not perform a pilot study, so we calculated the power of
the study post-hoc (page 10, line 14-15).
---
Referee 1 comment 11: Methods 8. What was the time interval between pretest, intervention, and posttest?
Authors’ reply: Everything was performed consecutively on one afternoon. We added the information in the methods section (page 7, line 19).
---
Referee 1 comment 12: Methods 9. Was reliability on questions, conversations, hands-on time, and tutor interventions also assessed?
Authors’ reply: We assessed tutor interventions per student, hands-on time, unrelated conversations and questions. These data are shown in figure 3. However, we did not record whether the same question or the same interventions occurred several times during the teaching session.
---
Referee 1 comment 13: Results 1. I notice that no confidence intervals were provided.
Authors’ reply: We decided to present most of our results as median and interquartile ratios, since the data were not normally distributed.
---
Referee 1 comment 14: Discussion 1. I encourage the authors to put more emphasis on their target group as a serious limitation to their work.
Authors’ reply: This is an important aspect, and we stressed it further in our revised manuscript (page 12, lines 19-23)

Minor Essential Revisions
Referee 1 comment 15: 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).
Authors’ reply: We now give a definition of simulation training and BLS at the beginning of the revised methods section (page 6, line 13-14)
---
Referee 1 comment 16: Methods 1. Please delete ‘one-way analysis of variance’ after ‘Kruskal-Wallis’
Authors’ reply: ‘one-way analysis of variance’ was replaced by “analysis” (page 9, line 14)
---
Referee 1 comment 17: Methods 2. Statistical analysis, 2nd paragraph: not sure what is meant by ‘irr’ here. Interrater reliability?
Authors’ reply: “irr” is the name of the extension package for the R software, in order to calculate the inter-class correlation. We added the information (page 9, line 24ff)
Referee 1 comment 18: Methods 3. Is a question/answer dialog strictly between tutor and trainee and a conversation only between trainees?
Authors’ reply: Yes, we defined a question/answer dialog for our raters as “a dialog between tutor and trainee”. It consists of a question from the trainee to the tutor and its matching answer by the tutor. Every question is counted. We added this information in the revised methods section (page 8, line 23ff)

Referee 1 comment 19: Methods 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’)?
Authors’ reply: We tested this hypothesis, there was no significant difference in the OSCE score after the training of the tutees of the individual tutors. The tutors adhered quite strictly to the teaching plan.

Referee 1 comment 20: Methods 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?
Authors’ reply: The training was intended to represent a “normal” small-group teaching session, and the available rooms did not allow the students to “hide”. Although the tutors were instructed to focus on the trainee, as with a normal training session, they could observe the other students as well. However, there was no specific instruction for the tutors to tell the other students that they were expected to observe.

Referee 1 comment 21: Methods 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?
Authors’ reply: That is correct. We further realized that we included 7 minutes in the original manuscripts, but the correct duration is 6 minutes. We corrected the manuscript accordingly.

Referee 1 comment 22: Methods 7. Who read the instructions to the students for BLS assessment?
Authors’ reply: The instructions were read to the students by the respective tutors. Tutors were however asked to only read the instructions, and questions addressing anything but comprehension issues were not allowed.

Referee 1 comment 23: Methods 8. We first read about self-assessment in the methods. The authors did not provide a rationale or literature about this in the intro.
Authors’ reply: We collected data on the self-assessment in order to measure the
Hawthorne effect. We expected the students in the smaller groups to feel better cared of although they received identical teaching. We modified the methods section (page 9, line 5-8) and the discussion (page 14, line 7-9) in order to stress this point.

---

Referee 1 comment 24: 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?

Authors’ reply: We decided to perform a baseline-test since the correlation between self-assessment and practical abilities is rather poor. The correlation between theoretical knowledge and practical abilities is better, but probably not good enough to exclude a sampling bias. Our priority was to exclude a sampling bias, which in our opinion would have hampered the quality of the data much more than the testing bias. As seen in the OSCE pass rates, only a minority of our students passed in the pre-test, while the majority passed in the post-test. This is a rather large effect for a testing bias. We stress this point in the revised discussion (page 13, line 3-6)

---

Referee 1 comment 25: Results 2. I am not really convinced of the added value of conducting a pre self-assessment questionnaire to students.

Authors’ reply: The self-assessment aimed at measuring the Hawthorne- effect. We expected the students in larger groups to feel less intensively cared of and thus not having learned as much as the students in the smaller groups although they received identical teaching. Indeed, the students in the larger groups were convinced to have learned less than the students in the smaller groups although that was objectively not the case. In even larger groups this attitude might actually affect the performance. We explain the rationale for the self-assessment in the revised methods section (page 9, line 5-8) and the discussion (page 14, line 7-9)

---

Referee 1 comment 26: Results 3. As an educator, I am interested whether student scores within a group were correlated. Did you take look at this?

Authors’ reply: We tested the hypothesis for students taught by the individual tutors (data not shown) but we could not show differences between the groups. We tried to control the sampling bias and the tutor effect by strictly randomizing the students and by the rigid teaching plan for the tutors.

---

Referee 1 comment 27: Results 4. Is there other published research applying the 42 points pass rate? Or perhaps this is incorporated in the Brennan checklist itself?

Authors’ reply: The checklist comprises 14 items, each rated from 1 (not fulfilled) to 4 (excellent). To our knowledge, there is no literature available covering a pass
level at the Brennan checklist. Therefore, prior to the study, we defined that – in order to pass – a participant had to earn at least 3 out of 4 in total (not for each item), which results in the 42-point pass rate.

---

Referee 1 comment 28: Results 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.

Authors’ reply: We arranged the data as suggested (page 9, line 10-15)

---

Referee 1 comment 29: Results 6. In table 1 we read ‘months since last resuscitation training’. This was not mentioned in the text, I believe.

Authors’ reply: We added the baseline characteristics in the revised methods section (page 8, line 8-10)

---

Referee 1 comment 30: 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.

Authors’ reply: We agree that even unrelated conversation can be positive in order to improve group dynamics and question concerning the topic can be negative (i.e. annoying or embarrassing). However, we did not perform an in-depth analysis on the quality of the comments. The raters were only asked to count the numbers of related and unrelated conversation. While this is a very interesting question, in our opinion one should employ a different study design in order to investigate the value of different comments. We amended the discussion (page 13, line 23-25).

---

Referee 1 comment 31: Discussion 2. Please use BLS abbreviated (last sentence 2nd paragraph) throughout the manuscript after it was written in full the first time you mention it.

Authors’ reply: The manuscript was adapted accordingly.

---

Referee 1 comment 32: 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)?

Authors’ reply: We revised the section. (page 15, line 13-19)

Discretionary Revisions

Referee 1 comment 33: Figures 1. Not sure what happened with Figure 1 but its
size is problematic (possibly due to upload/download).

Authors’ reply: The figure was graphically revised and provided as PDF and PNG file.

Referee 2

Major Compulsory Revisions

Referee 2 comment 34: 1. The authors have studied 4th year medicine students that were previously trained in CPR. It is evident from Table 2 that the objective CPR quality before the intervention was already good in all groups (with the exception of a lower compression depth in the X5 group). The authors admit that in the discussion, and they write in the limitations that these objective parameters are important and deserve more attention. I agree. Therefore, I see no proof that the intervention has had ANY effect on the objective parameters. That makes any further conclusion of differences between groups after training irrelevant. I therefore would challenge the conclusions that “group size does not influence the teaching success of BLS skills” and that “it is reasonable to teach CPR skills in larger groups”.

Authors’ reply: We agree that our study sample had a very good baseline skill, although the majority would have failed the resuscitation test. In your subsequent comment (35), you proposed to present the objective resuscitation data as % of students within the recommended ranges. When planning the study, we decided to present the data as proposed by Kramer-Johansen et al. However, we feel that your proposed way to present the data is a very good supplementation. We therefore analyzed the percentage of students within the ERC recommended compression ranges and extended table 2 accordingly. Although those resuscitation parameters were only secondary outcomes, this presentations now shows that the intervention also had a positive effect on the chest compression skills.

The conclusion was rephrased (page 15, line 12-19).

---

Referee 2 comment 35: 2. The authors report mean values of CPR quality in groups. If I understand correctly, these mean values are composed of mean values of individual performances. That is not a good presentation of the data. It would be much better to define what constitutes a pass score for the objective measurements, and to report the proportion of students in each group that reached the pass score. It may well be that this way of presenting the data reveals new insights in the differences before and after training.

Authors’ reply: We rearranged the data in Table 2 as suggested. Please see our reply to your comment 34 above for further details.

---

Minor Essential Revisions

Referee 2 comment 36: 1. It is unclear in the abstract that video recording was used. It is also unclear in the abstract what is meant by “teaching parameters”.

---
would also like to see in the abstract how data are presented (means of means, but I would recommend proportions of students passing a predefined level). The conclusions of the abstract are unclear. As I read the results, my take home message was that larger groups have less hands-on time and lower self-confidence. I would be careful with the conclusion that larger groups did equally well compared to small groups on the technical skills after training, because most students appeared to be already proficient before the training.

Authors’ reply: We revised the abstract in order to clarify these points.

---

Referee 2 comment 37: 2. I would suggest adding to the abstract that the duration of the training was adapted to group size (7 minutes training time per student). It is essential to the methodology. The fact that these 7 minutes were better used in the smaller groups, is an important finding.

Authors’ reply: We definitely agree and rephrased the abstract in order to clarify this point. As stated above, we realized that we specified the wrong duration in the original manuscript. Instead of 7 minutes per student, we had 6 minutes per student. We corrected the manuscript accordingly.

---

Referee 2 comment 38: 3. According to the methods, a group of 8 students would allow for a total training duration of 56 minutes. You showed that effective individual hands-on time was only 2 minutes in the groups of 8 students (for good reasons: more talking and maybe a sense of pressure). Given the fact that I have not seen proof in your data that the training actually resulted in an improvement of skills, I would conclude that the remaining 54 minutes were largely unproductive. I therefore have difficulty accepting the conclusion that larger groups are a good investment.

Authors’ reply: In the revised results section we display the data as pass rates, so the improvement in resuscitation skills is clearer. We rephrased the discussion (page 12, line 17-23) and the conclusion (page 15, line 12-19), since we definitely agree that groups of three or five students are preferable to groups of eight. In groups of eight we could detect a substantial decline of teaching quality and that we failed to detect a loss of skill is almost certainly due to the ceiling effect.

Discretionary Revisions
Referee 2 comment 39: 1. Please use “assessment” in stead of “diagnostics” throughout the manuscript.

Authors’ reply: The manuscript was adapted accordingly.

---

Referee 2 comment 40: 2. Please number pages when submitting a manuscript.
Authors’ reply: We added page numbering.

---

Referee 2 comment 41: 3. Under Outcomes: spent (spelling error)
Authors’ reply: We corrected the spelling error.

---

Referee 2 comment 42: 4. Under statistical analysis: statistically significant (spelling error)
Authors’ reply: We corrected the spelling error

---

Referee 2 comment 43: 5. In the abstract: chest compressions (no hyphen)
Authors’ reply: The abstract was corrected.

---

Referee 2 comment 44: 6. In the abstract: pass levels (no hyphen)
Authors’ reply: The abstract was corrected.

---

Referee 2 comment 45: 7. Table 1: proportion of males is redundant
Authors’ reply: We deleted the redundant information

Referee 3

Referee 3 comment 46: Regarding refusing 36 students to participate the study, all or some of these students might be low interest for participating in training courses therefore, it could bias in evaluation of differences between small and large study groups. It would be better to state it as a limitation.
Authors’ reply: We added the information in the discussion (page 15, line 5-7)

---

Referee 3 comment 47: Regarding the authors have mentioned the OSCE score in post test were above 90% in the study groups; Therefore, it would be difficult to evaluate the differences between groups accurately. It would be suggested the future studies design to assess between low interest students or with more complex content courses to find differences more accurately.
Authors’ reply: We agree and discuss the factors of teaching a less advanced audience or a more complex skill more extensively (page 12, line 17-23)

We hope that the revised manuscript is acceptable for publication and hope to be hearing from you soon.

Yours sincerely,
Moritz Mahling Anne Herrmann-Werner Nora Celebi