**Author’s response to reviews**

**Title:** Implementation of a large-scale simulation-based cardiovascular clinical examination course for undergraduate medical students – a pilot study

**Authors:**

Dimitri Arangalage (dimitri.arangalage@gmail.com)

Jérémie Abtan (jeremie.abtan@aphp.fr)

Jean Gaschignard (jean.gaschignard@aphp.fr)

Pierre-François Ceccaldi (pierre-francois.ceccaldi@aphp.fr)

Sid-Ahmed Remini (sid-ahmed.remini@univ-paris-diderot.fr)

Isabelle Etienne (isabelle.etienne@univ-paris-diderot.fr)

Philippe Ruszniewski (philippe.ruszniewski@aphp.fr)

Patrick Plaisance (patrick.plaisance@aphp.fr)

Victoire De Lastours (victoire.de-lastours@aphp.fr)

Agnès Lefort (agnes.lefort@aphp.fr)

Albert Faye (albert.faye@aphp.fr)

**Version:** 1  **Date:** 01 Aug 2019

**Author’s response to reviews:**

Dear Editor,

We were pleased to have the opportunity to revise our manuscript now entitled "Implementation of a large-scale simulation-based cardiovascular clinical examination course for undergraduate medical students – a pilot study".

The comments made by the Reviewers have greatly improved our manuscript, and we are very appreciative of such constructive feedback on our original submission. After addressing the issues raised, we feel the quality of the paper is much improved.
We provided point-by-point responses to each comment of the Reviewers indicating the changes we have made and where in the text these are to be found.

We do hope that you will find the revised manuscript acceptable for publication in BMC Medical Education and are looking forward to answering any questions that you may have.

Sincerely,

Dimitri Arangalage

RESPONSE TO THE REVIEWERS

Reviewer #1:

Asha N. Shenoi (Reviewer 1): The authors have reported the findings of a large scale simulation based cardiovascular semiology teaching programs for undergraduate medical students in an academic canter in France. Cardiovascular semiology is an important and topic and skills covered (cardiac auscultation skills, Peripheral Blood pressure measurements, and semiology of heart failure) are essential skills that every medical student should acquire regardless of the specialty they chose to pursue, so I really appreciate author's efforts in designing and conducting a study on this.

This study used a surveys to collect feedback on the student's acquisition of skills and student's and educator's perception of evaluation tools. They have reported the findings from the simulation course over 2 years and compared the results. Authors have concluded that implementation of a larger scale simulation program was feasible and students' perception of the course was favorable.

1. The strength of the manuscript is that it addresses an important topic. There is dearth of studies on feasibility of implementation of large scale simulation programs

We thank the Reviewer for this positive comment. We agree that studies on large scale simulation programs have seldom been reported in the literature.

2. This manuscript was easy to read. There are some minor spelling, spacing and grammar errors. Please correct them.
We modified the text as suggested by the Reviewer (line 110 and line 195).

3. Abstract - the conclusion part could be improved. In addition to mentioning that this large scale interventions are feasible, please consider adding a sentence describing the specific findings from the study. Under Results, line 56 is confusing. Please clarify half day sessions. Did authors mean two 75 minute sessions for each group and a total of 7 sessions to cover all the learners?

We agree with the Reviewer and improved the conclusion of the abstract by adding additional data: “Despite the need to commit significant human and material resources, the implementation of this large-scale program involving >400 students/year was feasible, and students and educators reacted favorably.” (line 65)

We also agree that “half day” is confusing and consequently removed it from the abstract and the text and replaced it with a precise description of the time spent at the simulation platform: “The course was divided into 2 consecutive compulsory sessions of 75 minutes each” (line 96). “In order to cover all the learners, a duration of use of the simulation platform of 28 hours/year was required.” (line 178).

4. Line 84 and 85 needs a reference.

We added the following references (line 83).


As suggested by the reviewer we divided the sentence (line 86): “In this article, we seek to describe the implementation and demonstrate the feasibility of a large-scale, compulsory, simulation-based cardiovascular semiology course for undergraduate medical students. Our main objective was to assess whether level 1 of Kirkpatrick’s model (evaluation of the degree of favorable reactions to learning events by participants) [15,16] was reached by analyzing the perceptions of educators and undergraduate medical students.”

6. Is there a formal definition for "large scale simulation training programs?"

The Reviewer raises an interesting question. As research studies focusing on teaching through simulation to an entire batch of students have seldom been reported, there is no formal definition of large scale simulation training programs in the literature for the best of our knowledge.

7. Line 85- Authors state the "the objectives of this article". What about the specific objectives of their study? What was the primary and the only aim to assess if learners reached kevel 1 of Kirkpatrick's model? Please specify.

We agree with the Reviewer and now clearly specify our primary objective: “Our main objective was to assess whether level 1 of Kirkpatrick’s model (evaluation of the degree of favorable reactions to learning events by participants) [15,16] was reached by analyzing the perceptions of educators and undergraduate medical students.” (line 88)

8. Line 95- Please change 1h15 to 75 minutes.

We modified the text as suggested: “The course was divided into 2 consecutive compulsory sessions of 75 minutes each” (line 96).
9. Line 104- what was the make of the mannequins (manufacturer details)

We now provide the name of the manufacturer: “Lifeform® Auscultation Trainer and Smartscope®” (line 106).

10. Line 113-'Authors mention 'considering communication with the patient in the scenario"

Did the learners use an actual patient? Authors describe later that they practiced the session 2 skills on each other, so please clarify this.

The learners did not use an actual patient but practiced skills on each other. We have now clarified this point in the text: “The second session was aimed at teaching blood pressure measurement using a manual sphygmanometer (Welch Allyn®), peripheral arterial auscultation as well as pulse localization and palpation, and finally the clinical examination of heart failure in a technical skill-based manner and considering communication with other students acting as simulated patients in a scenario.” (line 113)

11. Line 147-149 -Please explain why only one question was included in educator's survey

Only one question was included as we initially decided to mainly focus on students’ opinion. We agree that it represents a limitation of the study. We now clearly state this point in the limitations paragraph of the discussion: “Second, educators’ opinion evaluation was based on a single question.” (line 278)

12. Line 166- Please change 7 half day sessions to two 75 minute sessions' to keep the script uniform throughout the manuscript

Half day session no longer appears in the manuscript and has been replaced with two 75 minute sessions’. (line 96 and 133).

13. Line 170- Authors mention that the training was mandatory. Was he survey mandatory as well? How did they ensure high response rate (very unusual rate in voluntary surveys)? Was there multiple reminders and follow up needed to ensure response? Was the survey administered on paper or electronically?

The survey was also mandatory and had to be filled at the end of the session. It could be completed on a printed document or electronically. The following sentence now details this particular point: “The questionnaire was approved by the pedagogical committee of Paris-
Diderot University, mandatory, anonymous and could be completed on a printed document or on any mobile device.” (line 145).

14. Line 196- How were the short scenarios developed? Were these scenarios vetted by a team of experts?

The scenarios were developed by a team of expert cardiologists involved in the course. A panel of medical professors, members of the pedagogical committee of the Paris-Diderot University, reviewed and approved the scenarios. The following sentence has been added to the text: “All scenarios were developed by cardiologists and approved by expert members of the pedagogical committee.” (line 112).

15. Please include the manufacture details of the new mannequin the authors used - Were functionalities much different than the first one used. This could heavily influence the results.

We agree with the Reviewer and now provide the manufacture details of the new mannequin and describe the functionalities: “As a consequence, they were completely replaced and refunded the following year and an alternative supplier was used (SAM Basic® with a SimScope® stethoscope, Cardionics, Texas, USA). The new mannequins were tested before the purchase and had the advantage of presenting a wider range of customizable heart sounds of which the characteristics were much closer to reality, leading to the significant improvement of the feedback collected.” (Line 224).

16. Line 239-243. How can authors account for the effect of timing of the training on the result midyear vs beginning of the year? Please include this in limitations of the study.

We agree with the Reviewer and have added the following sentence to the limitations section of the study: “Third, changing the timing of the sessions, from mid-year to the beginning of the year, may have positively impacted the results of the surveys.” (line 278).

Reviewer #2:

Torben Wisborg (Reviewer 2): This manuscript concerns students' reaction to a cardiovascular teaching programme after introduction of mannequins for training auscultation and blood pressure measurement.
1. The authors deliberately sought after Kirkpatrick level one assessment only (students' reactions to the educational activity). This precludes outcome in form of actual learning. It is hard to believe that students would react negative to a new educational activity, especially when presented by teachers probably enthusiastic about their invention. This indicates a high risk of bias, and should encourage to the use of a strict methodological approach aiming to reduce this confounding. The background does not sufficiently explain the rationale behind the development of the new educational approach, nor does it describe the previous teaching methods. The aim is well described. We agree with the Reviewer. We would like to emphasize that this is a pilot study, and as a primary step in evaluating this new teaching method, we deliberately decided to assess whether Kirkpatrick’s level 1 was reached. We have added the following sentence to the limitations section of the study: “Moreover, the implementation of a new educational activity may be associated with increased enthusiasm from both educators and students, and may also have impacted the results of the surveys.” (line 280). We also modified the title of the manuscript by adding “– a pilot study” (line 2).

2. The methods does not describe the educational activity in sufficient detail to allow other to replicate it. In general, the manuscript would improve by adherence to the CONSORT & STROBE extensions (Cheng A et al. Advances in Simulation 2016; 1: 25). That said, being a deliberate intervention, the study ought to have been registered in a trials registry. I found no such registration described in the manuscript.

The present research work is only based on survey analysis and as a consequence it was not registered as a trial. The authorization from the pedagogical committee of the University of Paris to perform this research can be provided upon request.

3. The authors construct a composite result of 6 individual questions, but the exact calculation of this composite is not detailed to allow readers to understand the results.

As suggested by the Reviewer we have now added the following sentence to the statistical analysis section: “The total score presented in the students' survey is defined by the average (mean ± standard deviation) of the 6 individual questions” (line 163)

4. The results reports an impressive response rate. Free text fields are reported in a numerical fashion, although the calculations are left to the reader to guess.
As suggested by the Reviewer we have now added the following sentence to the statistical analysis section: “The analysis of the free response area is presented as the percentage of students expressing the same opinion. The free response areas of the questionnaire were independently analyzed by 2 independent observers, and opinions representing more than 20% of the study population were reported.” (line 164)

5. The discussion is extensive, and contains new information not given previously in the manuscript. This concerns scheduling of the teaching, change of simulators etc.

We have presented all quantitative data in the results section. A few qualitative data are developed in the discussion to facilitate the comprehension of the paper and enhance the interest of the reader.

6. I have some main concerns about this manuscript: The methodology is inappropriate. This is an attempt to do a controlled intervention. You would need a comparator, either a historical control or two parallel groups - or even the two years as compared to each other.

We would like to emphasize that this study is not an attempt to do a controlled intervention from one year to the next but to report the difficulties we encountered during the implementation of our education program. Our main goal was to evaluate whether Kirkpatrick’s level 1 was reached. We agree with the Reviewer that this particular point may be confusing for the reader. We have therefore clarified the main objective of our study: “In this article, we seek to describe the implementation and demonstrate the feasibility of a large-scale, compulsory, simulation-based cardiovascular diagnostics course for undergraduate medical students. Our main objective was to assess whether level 1 of Kirkpatrick’s model (evaluation of the degree of favorable reactions to learning events by participants) [15,16] was reached by analyzing the perceptions of educators and undergraduate medical students.” (line 86).

7. The intervention must be well described. It is difficult to follow the intervention: Which simulators were used for cardiac auscultation in the first and second year?

We now provide the name of the manufacturers:

The first year we used the “Lifeform® Auscultation Trainer and Smartscope®” (line 106).

“They were completely replaced the following year and an alternative supplier was used (SAM Basic® with a SimScope® stethoscope, Cardionics, Texas, USA). (line 224).
8. What was the blood pressure measurement experiment? Any changes here from year 1 to year 2? Equipment used?

Students were provided with a manual sphygmomanometer and a stethoscope to learn how to measure blood pressure on each other. The nature of the intervention during session 2 is now clarified: “students acting as simulated patients” (line 117).

There were no changes from one year to the next. We used Welch Allyn® sphygmomanometers. The text has been modified accordingly: “The second session was aimed at teaching blood pressure measurement using a manual sphygmomanometer (Welch Allyn®)” (line 113).

9. Kirkpatrick level 1 improvements are easily achieved, and the value of a such outcome must be substantiated by a meticulous description of the methodology used and the changes implemented. This would enable the reader to assess whether to try implementing the same changes into their curriculum. As the manuscript stands now it does in my view only report that pre-clinical students are happy when being taught clinical skills, and some (unreported) cardiac sound simulators seems to achieve better evaluation than others.

The manuscript has been modified and the methodology section, including the statistics paragraph has been enhanced.

- “The course was divided into 2 consecutive compulsory sessions of 75 minutes each aiming at teaching basic cardiovascular clinical examination skills to just over 400 students and took place at the iLumens Paris-Diderot platform (Université de Paris, Paris, France) dedicated to teaching through simulation.” (line 96)

- “The objectives of the first session were to teach the examination and palpation of the precordium, followed by heart auscultation on mannequins using an electronic stethoscope, including normal and pathological heart sounds (Lifeform® Auscultation Trainer and Smartscope®).” (line 103)

- “These scenarios consisted of brief and typical case presentations after which students had to use the stethoscope on mannequins to diagnose and characterize valvular conditions. All scenarios were developed by cardiologists and approved by expert members of the pedagogical committee.” (line 110).

- “All educators were junior staff attendings from various medical specialties, including non-cardiologists, provided with a detailed instruction manual describing how to use the mannequins, the educational objectives, and the duration of each step of the course.” (line 118).

- “A total of 8 mannequins dedicated to cardiac auscultation were used, as half of the students attended the first session involving auscultation of mannequins, and the second half
simultaneously attended the second session that did not require mannequins dedicated to auscultation. After 75 minutes, all students and teachers changed rooms to complete the other session.” (line 130).

As suggested by the Reviewer, we now provide the details regarding the cardiac simulators (line 106 and line 225).

Reviewer #3:

Yesim Senol (Reviewer 3):

1. Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached) The Introduction is poor. To start an article with a one paragraph quote from another source is odd. Far better that the authors start with some sentences of their own and then perhaps make a brief reference to this source. The article is more like the introduction of the program and This is not a research article

Level of interest: An article of limited interest.

We thank the Reviewer for his comments towards improving our manuscript. As wisely suggested by the Reviewer we have modified and improved the introduction as follows “The first year of medical studies in France is exclusively theoretical and students reach a real patient’s bedside during the second year. At our university, the teaching of clinical examination to second-year students was previously exclusively based on theoretical lectures and short bedside teaching sessions of 3 hours a week. As the time allocated to bedside teaching had markedly declined during the past years at our university, as well as in most institutions worldwide [1–4], the development of alternative teaching methods had become an essential task. As a consequence, we implemented simulation-based education program and a dedicated simulation center was created with the purpose of promoting simulation in healthcare education.

Simulation-based medical teaching has rapidly expanded over the past decade in the light of several studies that have consistently demonstrated its effectiveness in improving students’ skills and performance [5–12]. Although several studies have shown the benefits of such educational programs, large-scale implementations have seldom been reported [6–8,13,14], mainly because such programs represent a considerable investment, both in terms of financial and human resources.” (line 71).

We also agree that in the initial version of the manuscript the objectives were not precisely presented in the introduction. Consequently, we have now clarified the main objective of our study: “In this article, we seek to describe the implementation and demonstrate the feasibility of a large-scale, compulsory, simulation-based cardiovascular semiology course for undergraduate medical students. Our main objective was to assess whether level 1 of Kirkpatrick’s model
Reviewer #4:

Daisuke Son, M.D., M.H.P.E., Ph.D. (Reviewer 4): Thank you for submitting an interesting study. This is a research implementing and evaluating large-scale simulation-based cardiovascular semiology course without control.

Please consider revising the points below.

(Major points)

1. In background section, research question is not clear. If it's only "large-scale implementations were seldom reported", why such large-scale simulation is difficult to implement? What factors are barriers? Large-scale simulation platform is hardly available? Mannequin quality? The effectiveness is hampered easily with big-size group? Student-teacher interaction?

We agree with the Reviewer that the research objectives were not precisely presented in the introduction in the initial version of manuscript. We have now clarified the main objective of our study: “In this article, we seek to describe the implementation and demonstrate the feasibility of a large-scale, compulsory, simulation-based cardiovascular diagnostics course for undergraduate medical students. Our main objective was to assess whether level 1 of Kirkpatrick’s model (evaluation of the degree of favorable reactions to learning events by participants) [15,16] was reached by analyzing the perceptions of educators and undergraduate medical students.” (line 86).

We also improved the introduction as suggested by the Reviewer: “Although several studies have shown the benefits of such educational programs, large-scale implementations have seldom been reported [6–8,13,14], mainly because such programs represent a considerable investment, both in terms of financial and human resources” (line 82).

We would like to emphasize that the factors limiting the implementation of such programs have not been reported before. The quality of the mannequins (line 220) and the importance of the student-teacher interaction are developed in the discussion (line 243).

2. Kirkpatrick level 1 (satisfaction) evaluation is usually not acceptable with high-quality original research. Why authors did not conduct level 2 assessment? Please explain it. If it's only
level 1 evaluation, learning effect of this course is not validated. I recommend to add "---: a pilot study" at the end of the title.

We only conducted a level 1 assessment as this was a pilot study, performed concomitantly to the implementation of this new teaching program. This was a preliminary approach aimed at assessing students’ and educators’ satisfaction. This limitation is clearly presented in the limitations section “This study has several limitations [20]. First, the quality of the education program was assessed on the basis of the analysis of educators’ and students’ opinions, and students’ cardiovascular knowledge was not evaluated [21]. However, it is worth emphasizing that the objective of this pilot study was to assess whether level 1 was reached in Kirkpatrick’s model.” (line 274).

As suggested by the Reviewer, we have modified the title by adding at the end “a pilot study” (line 1)

3. In background (Page 5, Line 109), how many mannequins were used? We assume 16 mannequins for 16 simulation rooms. If authors replaced all mannequins next year, it must be very expensive. How were these costs covered?

We thank the Reviewer for raising this interesting question. We used 8 mannequins dedicated to auscultation training as half of the students attended the first session involving auscultation of mannequins, and the second half of the students simultaneously attended the second session that did not require mannequins dedicated to auscultation. After 75 minutes, all students and teachers changed rooms to complete the other session. The following sentence has been added to the text “A total of 8 mannequins dedicated to cardiac auscultation were used, as half of the students attended the first session involving auscultation of mannequins, and the second half simultaneously attended the second session that did not require mannequins dedicated to auscultation. After 75 minutes, all students and teachers changed rooms to complete the other session.” (line 130).

Following the negative feedback from students and teachers, the supplier accepted to refund the mannequins. The following sentence has been added: “As a consequence, they were completely replaced and refunded the following year and an alternative supplier was used (SAM Basic® with a SimScope® stethoscope, Cardionics, Texas, USA).” ( line 224).

4. In discussion (Page 11, Line 249), authors assume the high satisfaction was largely due to student-education interaction in a small-group. Please provide evidence for that (from the score or from description from the questionnaire). And how about that previously (bed-side teaching semiology)? Was it also small-group teaching?
As requested by the reviewer we now provide evidence from the questionnaire: “Finally, the availability of educators was appreciated (21%), as students highlighted the benefits of working in small groups (25%) that allowed questions to be asked more easily and facilitated educators-student as well as student-student interactivity, in a calmer setting than the hospital” (line 204).

Other factors also contributed to the high satisfaction rate as developed in the discussion “Indeed, most educators were very enthusiastic on the fact that they had a dedicated time for teaching with a restricted number of students, without having to manage multiple non-educational tasks concomitantly as it is often the case during traditional teaching sessions in hospital wards. Reciprocally, students appreciated the opportunity to benefit from the full attention of their educators and emphasized that they could ask questions freely. Thus, the proposed format encouraged the active involvement of all students through each step of the course, as they were constantly questioned on the cardiovascular clinical examination.” (line 243).

The bed-side teaching of semiology was also in small groups as now mentioned in the introduction “At our university, the teaching of clinical examination to second-year students was previously exclusively based on theoretical lectures and short bedside teaching sessions of 3 hours a week in small groups.” (line 72).

(Minor Points)

1. "Semiology course" is not understandable in English-speaking countries. "Diagnostics course" would be appropriate and easy to understand.

We have replaced “semiology course” with “diagnostic course” throughout the manuscript (lines 26, 47, 87, 95, 216)

2. Page 5, Line 95, "1h15" is "1h15m"?

We have replaced 1h15 with 75 minutes (line 96).

3. Please provide reference number of ethical approval (Declarations section, Page 14).

The letter from the pedagogical committee of the University of Paris approving the study can be provided upon request. This consideration is stated in the declarations section “The pedagogical committee of the University of Paris that deals with research authorizations and ethical considerations in the field of education has approved the study” (line 303).
Reviewer #5:

Jaime Yu, MD (Reviewer 5): First overall comment - although the paper adequately describes implementation of this new simulation curriculum to a large number of students, it does not provide adequate evidence that the educational outcome of this large scale and resource intensive endeavor is worth the major investments in infrastructure, finances, and time. The primary outcome of Kirkpatrick level 1, or essentially student satisfaction and teacher satisfaction, shows acceptability of use of simulation within the curriculum, but this is widely accepted and does not add to the current evidence regarding clinical skills teaching or use of simulation in medical education. It would have been very helpful, instead, if there was a more objective outcome measure to compare, such as performance on an end-of-training clinical skills examination, or some type of objective evaluation of student skills as a result of this educational modality.

Second overall comment - The use of the term "semiology" is not common in English language literature regarding this topic and is problematic. I would suggest the authors consider switching this to either "clinical examination" or "physical examination" teaching program, whichever they feel suits the objectives of the curriculum best.

We thank the Reviewer for his comments. We replaced semiology throughout the manuscript with either “diagnostics course” (lines 26, 47, 87, 95, 216) or “clinical examination” (lines 1, 51, 73, 97, 102, 115, 232, 237, 250).

We hope that after the major revisions we made, the manuscript will now be considered suitable for publication.

Specific manuscript comments:

1. line 95, methods - change 1h15 to 75 minutes

We modified the manuscript as suggested (line 96).

2. line 115, methods - What is the definition of "senior clinicians" who were the educators? Is this based on years out in practice? Rank in academic institution? Years of teaching experience? This needs to be clarified, and perhaps a bit of rationale as to why more senior rather than junior clinicians were chosen. With high clinician resource needs, many institutions may rely on a variety of clinicians to teach in clinical skills sessions, including resident physicians, junior staff attendings, senior staff attendings, etc.

We thank the Reviewer for raising this interesting issue. We agree that using “senior clinician”, as commonly used in France, is confusing. The correct definition is junior staff attendings. We have now modified the manuscript as suggested: “All educators were junior staff attendings from
various medical specialties, including non-cardiologists, provided with a detailed instruction manual describing how to use the mannequins, the educational objectives, and the duration of each step of the course.” (line 118).

3. line 154, methods - authors specify that they used either Wilcoxon test or chi-squared "as appropriate", and then later in results it does not specify which test provides which p-value. Strongly suggest this is clarified - which data is subjected to which test? What significance is being reported

We have now clarified the statistics section “The Shapiro-Wilk test was used to evaluate distribution among variables. As continuous variables were not normally distributed in the present study, the Wilcoxon test was used for comparison between groups. The χ² test was used for comparison between categorical variables.” (line 160).

“A p value < 0.05 was considered statistically significant.” (line 168).

4. line 174, results - "Students' overall appreciation improved" - which measure is the "overall appreciation" - this is not listed as a separate variable in the methods or in table 1. Need to specify - is this the mean score or sum score across the other 6 measures? How is this value determined? Also, this indicates change between the 2016-2017 cohort and the 2017-2018 cohort. Need to specify what is different between the two groups and why improvement occurred? This was not clear at this stage of the paper. The specific results listed next in lines 176 through 178 are subject to same question. "appreciation" for session 1 and 2 by students is listed, but yet this is not listed in methods or in table 1 as a specific outcome variable. Also, not clear then what test would be used to provide the significance listed in line 177 for the between group differences noted

The overall appreciation refers to the total score in table 1. To clarify this point table 1 was modified and now clearly presents the total score as the overall appreciation (page 20). As suggested by the Reviewer we have clarified the calculation of this score by adding the following sentence to the statistical analysis section: “The total score presented in the students’ survey is defined by the average (mean ± standard deviation) of the 6 individual questions” (line 163).

The results section provides quantitative results on students’ and educators’ opinions and these results are discussed in the discussion section. The explanation on the improvement is provided in the following paragraphs: “A major issue encountered during the first year of implementation was related to poor quality simulation mannequins, which represented by far the main complaint by students and educators. Mannequins were considered of poor quality, with heart sounds different from reality and not in line with the national curriculum and the theoretical objectives
of the local second-year program. As a consequence, they were completely replaced and refunded the following year and an alternative supplier was used (SAM Basic® with a SimScope® stethoscope, Cardionics, Texas, USA).” (line 220); “During the first year, the teaching sessions were scheduled mid-year, after the students had started learning the clinical examination by examining real patients in wards. Although the appreciation of students was overall positive, many of them mentioned that the simulation-based teaching program would have been much more useful if it had been scheduled before reaching a real patient’s bedside. Consequently, the sessions were scheduled at the beginning of the following year.” (line 231).

A chi-square test was used for comparison between groups as now clarified in the methods paragraph: “The $\chi^2$ test was used for comparison between categorical variables.” (line 162).

5. lines 182-190, results - this section discussed a logistic regression. However, I am not clear what the question is that the regression is supposed to answer? Does this section imply that quality of simulation equipment WAS related to all evaluation measures EXCEPT educator-student interactivity? Need to be more clear on why logistic regression was chosen and what problem this was meant to solve

We agree with the Reviewer that the use of a logistic regression in the present study may be confusing for the reader and that it does not provide substantial additional value to the results as well as for the their interpretation. Consequently, we removed this analysis from the manuscript.

6. line 191, results - “the analysis of the free response area” - how did this analysis occur? Was this by pre-determined categories or themes? How many authors analysed the free responses? This is not described in the methods and should be, as it is an entirely separate analysis from the quantitative measures described thus far

We thank the Reviewer for his comment. We have now added the following sentence to the methods section: “The analysis of the free response area is presented as the percentage of students expressing the same opinion. The free response areas of the questionnaire were independently analyzed by 2 independent observers, and opinions representing more than 20% of the study population were reported.” (line 164).

7. line 210, discussion - "observed a significant improvement from the first to second year" - why is this change observed? what was different between the years, and why is this change the most significant finding of this program evaluation? There is some discussion later in the paper about change in timing of sessions, and it is repeated throughout that the simulators themselves
were changed. These details need to be part of the discussion earlier so that this finding can make sense.

We agree with the Reviewer and have now modified the discussion section. As suggested by the Reviewer all details explaining the causes of the change observed and the modifications we implemented are now presented earlier in the discussion (line 220 to 252).

8. line 218, discussion - "their learning considerably improved after training" - how was this shown? The paper discusses primarily student and educator satisfaction, and has some program evaluation information and shows a responsive program with changes to the simulators and changes to timing made. However, there are no objective measures in the paper to show that students improved in their knowledge or skills compared to prior to this program, or compared to prior years where simulation of this type was not used.

We agree with the Reviewer and have removed this part of the sentence from the manuscript. “Despite the need to commit significant human and material resources to carry out this educational project, students reacted favorably to the training sessions.” (line 256).

Reviewer #6:

Marko Zdravkovic, M.D. (Reviewer 6): General comments:

I would like to commend the authors on implementation of the simulation course for their students as well as the improvements they managed to achieve from the first to the second year. Having said that, my major concern with this manuscript is that it can be hardly justified to publish as original research article. A more appropriate format would be a short report on an educational intervention and its evaluation.

Overall, one way to decrease the faculty burden is to include near peer-teaching (senior students teaching younger students). This would also improve students' teaching skills as future doctors are expected to be able to teach the new generations of physicians.

Specific comments on the manuscript improvement:

Background:

1. I would omit the word semiology - it is a study of signs. I have never encountered this word in medical studies before (despite being trained in several European countries) and it might be more often used in French speaking environments. I doubt many readers will understand what you
mean by it. I would recommend to definitely remove it from the title and the manuscript, but if you want to keep it in the text, you should define what it means at the first instance.

We thank the Reviewer for his comments. We replaced semiology throughout the manuscript with either “diagnostics course” (lines 26, 47, 87, 95, 216) or “clinical examination” (lines 1, 51, 73, 97, 102, 115, 232, 237, 250).

2. p. 5, line 95: 1h15 / please correct: 75 minutes
We modified the manuscript as suggested (line 96).

3. p. 5, line 96: please be specific instead of > 400 students write the exact number or replace with: just over 400 students
We agree with the Reviewer and have modified the manuscript as suggested (line 97).

4. p.7 line 159: please omit reasonable. The data is either available upon request or not. And you should define what data is available rather than describe it as reasonable request.
As suggested by the Reviewer we removed “reasonable” (line 169 and 308).

Method:
5. This is more appropriate format for a short evaluation report rather than a full 3,000+ word article. Content validity and piloting of the questionnaires are not clear.
We hope that after the major revisions we performed the editors will now found our paper eligible for publication as a full paper.

Results:
6. Table 1: even though the statistical tests performed show significant difference, it is of little "value", for example quality of teaching method 2.60±0.57 vs 2.75±0.45 (p=0.001). This tends to happen with large samples. As appropriately reported in the text, Quality of the simulation equipment is the only one that truly stands out.
We thank the Reviewer for his comment. As mentioned, we have already reported this particular point in the text.

Discussion:

7. Overall, this paper is too long given what has been done. It is more of an evaluation of a course and should at most be published as a short report. It does not warrant to be regarded as an original research article.

We have now considerably modified the manuscript in the light of the Reviewers comments and hope that the editorial board will find the article suitable for publication as a full paper.

Conclusion:

8. Given what you studied, skills acquisition would be a more proper term rather than knowledge (spending as you report significant human and material resources for knowledge acquisition though simulation is not cost-effective). Furthermore, what behavioral changes do you expect with refereeing to them in the conclusion? I would suggest omitting this part.

We agree with the Reviewer and have modified the conclusions as suggested: “Further studies evaluating students’ skills acquisition following the teaching program are necessary to confirm the benefits of this method.” (line 293).

Declarations:

9. AVAILABILITY OF DATA AND MATERIEL - please redo spell check for the entire document; it is unclear to me why anonymity would be questionable here and a reason not to share the database - survey responses were anonymous or did students have to report their names on the forms?

We checked the spelling in the document. The survey was anonymous. Consequently the paragraph was modified: “AVAILABILITY OF DATA AND MATERIEL: The datasets generated and/or analyzed during the current study are available from the corresponding author on request.” (line 309).