PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf)** and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Introducing physically active lessons in UK secondary schools: feasibility study and pilot cluster-randomised controlled trial
AUTHORS	Gammon, Catherine; Morton, Katie; Atkin, Andrew; Corder, Kirsten; Daly-Smith, Andy; Quarmby, Thomas; Suhrcke, Marc; Turner, David; van Sluijs, Esther

VERSION 1 – REVIEW

REVIEWER	Nicole Nathan Hunter New England Population Health and The University of Newcastle, NSW Australia
REVIEW RETURNED	17-Jul-2018

GENERAL COMMENTS	Thank you for the opportunity to review this paper, it was a pleasure to read and is a good example of a pilot study. I have very few questions or suggested edits for the authors thus I hope my comments are useful and can be addressed within word limit restrictions. Abstract • The abstract is missing cost as an aim of the project. • Participants- what was the total number of eligible students/ teachers? As this goes to feasibility? Or was those that enrolled the total sample? Background • Line 93- would be useful to give some context of preliminary effectiveness of what i.e. student physical activity? Feasibility Study • Line 99- would be useful to be consistent in the order of the wording i.e. acceptability, cost and preliminary effectiveness- I think this has changed throughout (i.e. line 91-92 different order). • Line 118- attend • Line 126- might be useful to give a brief definition of “active pedagogical approaches” • Line 148- do you have any indication of what other costs to the school? i.e. teacher time to attend the workshop? Teacher time to make changes to their teaching practices in order to implement this? If not would be good to include this as a limitation (or future work needed) regarding calculating the costs to schools to implement such practices. • Line 173- this is the first time mental health and wellbeing introduced which was a bit of a surprise. If this is a measure of preliminary effectiveness maybe add into that definition earlier. It would be useful to have more detail around the measure e.g. how many items? What scale was used?

	• Similarly for time-on-task. This is more expected but be good to introduce it earlier. How many lessons were observed? • Analysis section is quite light on- for example need more detail regarding how mental wellbeing and on-task behaviour was scored and analysed. • Line 193- sentence begins with a number • Line 93- what was the total potential sample- good to know for feasibility. • Line 198- 11 teachers attended both sessions- out of how many potential teachers? Again goes to feasibility and acceptability. • Line 232- given this is an aim of the study think that more detail should be given about the preliminary effectiveness for activity, mental health and on-task behaviour rather than just referring the table. Using the headings as presented in the measures rather than "student measures" would also be helpful- unless it is edited throughout to say preliminary effectiveness on student outcomes i.e. activity, mental wellbeing and on-task behaviour. Pilot study • Line 303- seems a bit odd sitting here. As you have already referred readers to Table 1 can you simply include this detail there? Initially I thought that all the other measures had been dropped so it may also avoid that confusion too. Pilot study • Analysis section missing? • Preliminary effectiveness- as per my comments above it would strengthen the paper to have some intext results rather than directing to the table. • Line 420-426- if the effectiveness was also measured by wellbing and on-task need to make comment here on effectiveness on these outcomes. Reflections/ discussion and conclusion • Line 467- no effect on activity as well as wellbeing and on-task. • It is really great to see a study published that has had mixed results. Thanks to authors for sharing these learnings.

REVIEWER	Emma Norris University College London, UK
REVIEW RETURNED	27-Jul-2018

GENERAL COMMENTS	This paper presents two exploratory studies of a physically active lesson intervention in secondary schools This paper is valuable in its unique whole-school approach at secondary school level and its honest account of the difficulties in implementation and lack of outcome effects. It's lack of prescribed dose of active lessons is also interesting, reflecting the realistic ability of teachers to implement such lessons within their existing constraints. I have some suggestions for minor improvements: Introduction 1) Add reference to McMichan paper https://www.ncbi.nlm.nih.gov/pubmed/29570032 on secondary PA in classroom Methods 2) Evaluation of the pilot/feasibility studies may benefit from a framework such as RE-AIM, to help better structure the assessment of feasibility. 3) The Measurements section of the Feasibility study would benefit from being separated into evaluation vs outcomes. This would just require addition of some sub-headings.

	4) Add how long student questionnaires took to complete 5) Unclear when and for how long accelerometers were given to pupils: mentions they were told could wear for one week but unclear if scheduled e.g given on Mondays 6) Specify if any Behaviour Change Techniques https://www.ncbi.nlm.nih.gov/pubmed/23512568 were used in teacher training within the Methods e.g goal setting Results 7) Line 217-221: too much repetition of 'the assistant head teacher' 8) Line 353 – mouthguards and sport shoes purchased by pupils. It seems concerning that additional resources were being purchased by students. It might be worth elaborating on this as an issue in the Discussion.

REVIEWER	Doctor Brendon Hyndman Charles Sturt University, Australia
REVIEW RETURNED	13-Aug-2018

GENERAL COMMENTS	Overview The research team and project of study has clearly undergone rigorous review and the manuscript reflects extensive preparation, steps/considerations and the level of writing was first class. As a qualified teacher who has worked in schools and is now a teacher educator, I found some concerns with the approach to be able to work in with teachers that teach subjects that are not 'movement focused'. Many teachers who choose to teach in alternative discipline areas do so as they did not have very positive physical activity experiences during their primary/secondary schooling. The arguments from an epidemiological standpoint and the background work to promote physical activity in schools are clearly compelling, yet the disengagement from teachers suggests more could be considered from an 'educational' viewpoint. Even with the most soundly designed projects and rationale for interventions, it could take a lengthy time to modify some teachers' negative perceptions around physical activity who focus on other areas of schooling. By addressing a number of amendments in the manuscript, the study can serve to showcase a number of implementation considerations for researchers who are considering undertaking similar projects in the future. Specific considerations 1) Page 7, Line 155/Page 8, Lines 156-157. There are concerns in many education circles about the measurement of students' weight at schools being problematic (alongside BMI being an inaccurate measure). Were there procedures used to minimise these concerns around students having their body weight measured? This could be a valuable inclusion. 2) Page 8, Lines 158-164. The rationale for using the triaxial wrist-worn accelerometers was based around 'energy expenditure' (& participation compliance), yet there was no data based specifically on energy expenditure in the results sections. Does this mean that the accelerometer intensity data presented is not valid? This needs to be looked into.

	3) Page 9, Line 194. Again, think classifying as overweight/obese based upon BMI is problematic. Perhaps a limitation. 4) Page 9, Lines 197-198. Perhaps need to specify somewhere in the manuscript the implications of approximately 5 teachers not attending both sessions on the design/outcomes of the study. Similarly, later in the pilot study section (line 310). 5) Page 9, Lines 197-202. Statistics were mentioned of improvement, it would be worth clarifying if these improvements were 'statistically significant'. 6) Page 10, Lines 203-210. It appears as though there were a number of challenges for the teachers to deliver this type of intervention (see overall summary earlier in review). Many students in schools will also look for any opportunity to get out of 'English/Mathematics tasks/learning'. These training issues appear to be major limitations. It would be worthwhile trialling PAL in any non-movement subject areas in which the students genuinely want to participate and learn in. 7) Page 11, Line 238. It states 'teacher acceptability should be explored further', yet for the feasibility study to inform the pilot study, it would have been better from an educative perspective to focus on the training regime, preparation of both students and teachers. 8) Page 12, Lines 267-270. More detail into why it was not possible to blind is required. This is a major limitation. 9) Page 12, Paragraph 3. Please clarify if the schools were different in the pilot study to the feasibility study. This is important information to include. 10) Methods overall. It states 'full-scale randomised controlled trial'. Yet in the methods there are only two schools. Please justify this amount of schools from the literature for a randomised controlled trial. Was there a power analysis conducted? This all needs to be included to justify the approach to the research. Is full-scale necessary? 11) Level of teaching experience of the teachers is important in this section (& at the school being measured). This could have a major bearing on how open minded the teachers were to the PAL intervention and not being entrenched in particular approaches (non PA). 12) Page 15, Paragraph 2. From an educator perspective, it appears as though the students were taking advantage of the PAL intervention to escape learning. 13) Page 16. A really important methodological point to have considered is 'level of experience facilitating physical activity or physical education activities' etc. Not in terms of the PAL (specified on page 18, paragraph 1), yet in terms of their careers. Also, determining teachers' beliefs, experiences and intentions around physical activity prior to delivery. This could help plan for a lot of the barriers identified.

	14) Page 18, Paragraph 1. The 'acceptability of the training was not demonstrated in the feasibility study'. Please correct. 15) Page 18. Important consideration for implementation is 'to include teachers in developing the study/design that would be anti-PE or who normally avoid facilitating physical activity where possible.' Gaining such teachers' thoughts would be invaluable to implement such a project successfully. 16) Page 20. Strength was definitely the formative work for interventions. Well done. 17) Page 20. Perhaps some of my earlier points mentioned relating to limitations here. It still wasn't clear the relationship between rationale/reliability of the accelerometers (energy expenditure) and the results findings (various intensities). 18) Page 20, Conclusion, Line 462. Yes, but teacher acceptability was also 'not demonstrated'. This should be re-written. 19) Page 20, Conclusion, Line 464. Was the intervention really acceptable to the students? I think there could be more scrutiny here around students looking to escape learning practices (what are the implications of being behind in the curriculum?). Were there gaps between communication between school leaders and teachers? E.g. School leaders wanting to shine by aligning with a major health initiative, teachers wanting to shine to the school leader by participating, yet in reality the teachers perhaps not wanting their classrooms distracted by the physical activity integration? Could this have led to there being some positives about the PAL integration? Some considerations around the findings, processes here. 20) Page 24, References. Could consider research on teachers (& early teachers e.g. PSTs) in how they report on difficult schooling experiences related to physical activity and physical education (PE) as pursuing non-PE specialist teaching. Early PE experiences can be tough for some to overcome. John Haynes from UNE wrote a piece on this and there are some snippets within the GET-PE study. 21) As it stands, the title appears as though the school recruitment in the UK is more widespread than in reality. Authors could consider perhaps reframing the title to: 'Piloting physically active lessons in UK secondary schools...' or something similar to note the relatively small sample. Minor editing considerations 1) Line 27. Please insert 'was' after training. 2) Throughout manuscript please look into the use of apostrophe with students. e.g. Line 39 is students' rather than student's 3) Would be good to insert a few brief examples of programme components in brackets somewhere to showcase brief elements of what the physical activities entailed (this could help generate interest from physical activity professionals with specific interests). An example would be line 37 4) Line 193 and other incidences. Better to start sentence with N=99 than 99. Perhaps just refer to 92% on this line rather than another number.

	5) Line 194. Would be good to specify the standard deviation here for 13.0 years. 6) Page 10, Line 211. Perhaps just percentage here. 7) Line 306. Again, go with N= to start sentence and referring to 92% here will be fine without the brackets and referring to 205.

REVIEWER	Dr Ash Routen Loughborough University, UK.
REVIEW RETURNED	31-Aug-2018

GENERAL COMMENTS	Comments to author: This is an interesting and well-conducted piece of work, which reports on both a feasibility and pilot study of a secondary school classroom physical activity intervention. It is to my knowledge the first of its kind in secondary populations, and the author's should be commended for this and the effort required to collect this volume of varied mixed methods data in a challenging setting. I have a few minor comments that require clarification, please see below. P2 Line 34 and line 38: Need an adjective before acceptability e.g. poor, excellent etc. P4 Line 60 and line 63: Need space before citation brackets. Other occurrences in the manuscript need amending. P4 Line 62: Insert comma after 'interventions'. And remove comma after 'hours'. P4 Line 70: Should this instead read something like 'perceived to be the most-feasible', acceptable etc.? As to my knowledge this is the first PAL intervention in secondary settings, in the UK at least. P5 Line 90: It would be good to set out more clearly in this paragraph exactly what each study intended to do, rather than blending the aims. This is important because as you read on the purpose of the feasibility study isn't overt for the reader. P6 Line 111: Why 120 students? Was this a pragmatic choice due to resource/manpower available? If so it's worth stating this. Also why Year 7 and Year 9 students? P6 Line 120: Assume this is five PAL lessons? Suggest inserting 'PAL' to clarify. P6 Line 123: Where was the training delivered? What time in the school year? Did teachers have to be bought out of usual teaching time? A few of these details would be useful. P8 Line 177: I'm a little confused as to why time-on-task was included as a pre-post measure. Usually it is used to measure TOT during or post PAL. Did you expect TOT to change due to children having taken part in PAL? Also what was the protocol for the observation, did you for example focus on one table at a time and work around in a systematic order focusing on one student at a time?

	P11 Line 231: Effectiveness could be re-worded given that the study was not powered to detect change in these variables and due to a lack of control group (as you acknowledge). Page 12 Line 262: Why did you decide to recruit three schools and were they selected based on size, SES, ethnicity etc.? It would be helpful if some information on the schools were presented, for example what was their Ofsted rating, school size, SES etc. P16 Line 359: I think you should mention the issue of follow-up students being different to baseline students for some of the measures. Was this due to practical reasons? An interesting point re. feasibility of data collection in this population perhaps. P18 Line 392: Did the teacher's comment on the duration of the training and if they required a top-up, or if they felt competent enough to deliver PAL once trained? My experience in primary teachers is that even a full day of training is insufficient. I wonder therefore if 4 hours of training is sufficient to change teacher's pedagogy? It's quite a big step change for them, unlike for example integrating a few standing desks in the corner of the room. Table 1. Were the follow-up measures taken whilst the teacher's were still implementing PAL? As from table 3 it appears you have accelerometer data from PAL lessons, but I can't see where in the main text of the manuscript that you describe how you collected this data. This also relates to your point in the discussion, where I think you are right in stating that it might be a stretch for accelerometers to capture short periods of PAL lessons among the noise of an overall day. Therefore one limitation is that you weren't able to collect direct observation of what they were doing in a sample of the lessons, which might help explain the small dose of PA received.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 Comments

1. The abstract is missing cost as an aim of the project.

Cost has been added to the objectives section of the abstract.

2. Participants- what was the total number of eligible students/ teachers? As this goes to feasibility? Or was those that enrolled the total sample?

These details are provided in Supplementary Tables 1 and 2:
Supplementary Table 1: The total number of eligible students was 360.
Supplementary Table 2: The total number of eligible teachers was 15.

3. Line 93- would be useful to give some context of preliminary effectiveness of what i.e. student physical activity?

The following text has been added to clarify outcome variables that were measured to assess preliminary effectiveness of the intervention:

“The aim of the feasibility study was to assess (i) the feasibility, acceptability, costs, and preliminary effectiveness (for reducing sedentary time and improving wellbeing and time-on-task among students) of a PAL training programme for secondary school teachers, and (ii) the feasibility and acceptability of study procedures.”

4. Line 99- would be useful to be consistent in the order of the wording i.e. acceptability, cost and preliminary effectiveness- I think this has changed throughout (i.e. line 91-92 different order).

The wording order has been edited for consistency throughout the manuscript.

5. Line 118- attend

Corrected

6. Line 126- might be useful to give a brief definition of “active pedagogical approaches”

The following text has been added for clarification:

“The focus was on supporting teachers to adopt active pedagogical approaches (teaching strategies that incorporate activity), rather than providing new, PAL plans.”

7. Line 148- do you have any indication of what other costs to the school? i.e. teacher time to attend the workshop? Teacher time to make changes to their teaching practices in order to implement this? If not would be good to include this as a limitation (or future work needed) regarding calculating the costs to schools to implement such practices.

The training was delivered as part of ongoing teacher development and training attendance was therefore not costed. We have added the following text to the intervention section of the feasibility study methods:
“The training was delivered at the intervention school between March and April, during pre-scheduled after-school teacher-training time.”

We have also added the following text to the overall discussion: “Schools scheduled the PAL training during pre-scheduled after-school teacher-training slots, as such, the intervention did not require teachers to attend any more after-school training than they typically would.”

8. Line 173- this is the first time mental health and wellbeing introduced which was a bit of a surprise. If this is a measure of preliminary effectiveness maybe add into that definition earlier. It would be useful to have more detail around the measure e.g. how many items? What scale was used?

Text has been added to the feasibility study introduction to clarify the outcome variables that were used to assess preliminary effectiveness:

“The aim of the feasibility study was to assess (i) the feasibility, acceptability, costs, and preliminary effectiveness (for reducing sedentary time and improving wellbeing and time-on-task among students) of a PAL training programme for secondary school teachers, and (ii) the feasibility and acceptability of study procedures.”

In our efforts to keep the word count as low as possible, we have provided the reference for each measure that was used to assess mental health and wellbeing indicators in the feasibility study methods section. Information regarding number of items and scoring instructions can found in the cited references. We hope this is acceptable to the editor.

9. Similarly for time-on-task. This is more expected but be good to introduce it earlier. How many lessons were observed?

Text has been added to the feasibility study introduction to clarify the outcome variables that were used to assess preliminary effectiveness (please refer to our response to comment 8). The number of lessons observed has been added to the 'time-on-task' section of the feasibility study methods: "Students' time-on-task was assessed during three lessons by one member of the research team"

10. Analysis section is quite light on- for example need more detail regarding how mental wellbeing and on-task behaviour was scored and analysed.

As above, we have tried to keep the word count low. We have added the following text to the 'mental health and wellbeing' section of the feasibility study methods: "All questionnaires are validated for use with adolescents and were analysed according to published instructions (29,30,37)."

Further information on the scoring and analysis of on-task behavior has been added to the time on task section of the feasibility study methods:*"Time-on-task:* Students' time-on-task was assessed during three lessons by one member of the research team using a momentary time-sampling procedure (which incurs less bias than other sampling procedures[33, 34]). At the start of each observed class, the teacher asked all students participating in the study to raise their hands. From the students that raised their hands, the researcher identified two boys and two girls (when possible) to observe. The researcher chose students sitting in different areas of the classroom. Each student was observed once per minute, in a consistent order, for the duration of the lesson. Students' behaviour was coded as: (i) on-task, (ii) off-task-passive, (iii) off-task-motor, or (iv) off-task-noise[35]. The mean percentage of intervals recorded as 'on task' for observed students and classes was calculated and used as the outcome measure."

11. Line 193- sentence begins with a number

This has been adjusted.

12. Line 93- what was the total potential sample- good to know for feasibility.

This information has been added to the results section: Of 120 students invited to participate in the evaluation measures, 99 were recruited, with 91 (92%) providing data at baseline and follow-up.

The information is also provided in Supplementary Table 1.

13. Line 198- 11 teachers attended both sessions- out of how many potential teachers? Again goes to feasibility and acceptability.

This information has been added to the results section: Training session one was attended by 14 (out of 15) teachers

This information is also provided in Supplementary Table 2.

14. Line 232- given this is an aim of the study think that more detail should be given about the preliminary effectiveness for activity, mental health and on-task behaviour rather than just referring the table. Using the headings as presented in the measures rather than "student measures" would also be helpful- unless it is edited throughout to say preliminary effectiveness on student outcomes i.e. activity, mental wellbeing and on-task behaviour.

Preliminary effectiveness was one of the aims of the study, we have now added some additional text to

the results to supplement the reference to the table: “Table 2 summarises baseline and follow-up data for all student measures. Sedentary time increased by 8.7 minutes and time spent in light-intensity activity decreased by 8.1 minutes. Minimal changes were observed in the mental health and wellbeing scores between baseline and follow-up.”

15. Line 303- seems a bit odd sitting here. As you have already referred readers to Table 1 can you simply include this detail there? Initially I thought that all the other measures had been dropped so it may also avoid that confusion too.

The line of text that the reviewer refers to is providing clarification to the sentence “Data collection followed the same procedures as described for the feasibility study, except for the assessment of PAL dose and time on task” (written 5-7 lines above the text that the reviewer refers to).

We feel this line of text provides readers with clarity on how the procedure for assessing time on task was different in the pilot study. Following another reviewer’s comments, some information has been added after this line of text, which makes it a more substantial sentence:

“Time on task: Four lessons were observed at baseline and follow-up, at both schools. At baseline (prior to delivery of PAL training) the research team observed typical desk-based lessons. At follow-up, the research team asked to observe physically active lessons.”

16. Pilot study - Analysis section missing?

Apologies for this omission, a brief analysis section has been added.

17. Preliminary effectiveness- as per my comments above it would strengthen the paper to have some in text results rather than directing to the table.

In addition to referring to the table, we provide the key messages from the data in this section. This is consistent with the amount of detail provided in the feasibility study results section. We have revised the results section to include more description of the results: Table 3 presents activity intensity during PAL at follow-up and the equivalent lesson at baseline (excluding P.E. and drama lessons). There was no evidence of changes in sedentary activity or time spent in light, moderate and vigorous activity intensities. Table 4 summarises baseline and follow-up values for all outcome measures for intervention and control participants. There was no evidence of preliminary effectiveness on sedentary time or light activity, or on indicators of mental health and wellbeing (including academic efficacy, positive & negative affect, and disruptive behaviour).

18. Line 420-426- if the effectiveness was also measured by wellbeing and on-task need to make comment here on effectiveness on these outcomes.

We had added some detail for clarity: “There was no evidence of preliminary effectiveness on sedentary time or light activity, or on indicators of mental health and wellbeing (including academic efficacy, positive & negative affect, and disruptive behaviour).”

19. Line 467- no effect on activity as well as wellbeing and on-task.

Text has been added to address this comment.

Reviewer 2

1) Add reference to McMichan paper https://www.ncbi.nlm.nih.gov/pubmed/29570032 on secondary PA in classroom

Thank you for highlighting this, the reference has been added.

2) Evaluation of the pilot/feasibility studies may benefit from a framework such as RE-AIM, to help better structure the assessment of feasibility.

Thank you for this suggestion. We used the RE-AIM framework to support study design and data collection decisions, however, the aims of our studies are not completely aligned with the RE-AIM components. For example, understanding intervention and PAL feasibility and acceptability were of greater importance than establishing effectiveness and maintenance (at this feasibility/pilot testing stage). In addition, a focus of our study was on the feasibility and acceptability of study measures, which doesn't fit into the RE-AIM framework.

We feel we have clearly conveyed our research aims and that throughout the article we have systematically presented findings and reflections in relation to those aims. Unless the editor thinks that re-structuring the evaluation would substantially enhance a reader's ability to comprehend the conclusions, we would like to leave the structure of the discussion as it is.

3) The Measurements section of the Feasibility study would benefit from being separated into evaluation vs outcomes. This would just require addition of some sub-headings.

Sub-headings have been added as suggested.

4) Add how long student questionnaires took to complete

This detail has been added to Table 1: Questionnaire (15 minutes)

5) Unclear when and for how long accelerometers were given to pupils: mentions they were told could wear for one week but unclear if scheduled e.g given on Mondays

This detail has been added to the activity monitor section of the feasibility study methods:

Participants were given verbal and written instructions on monitor wear, including that the monitor was waterproof and could be worn continuously for the next seven days (Monday to Monday).

6) Specify if any Behaviour Change Techniques https://www.ncbi.nlm.nih.gov/pubmed/23512568 were used in teacher training within the Methods e.g goal setting

We have included the following information about the behaviour change techniques underpinning the training on pages 6 and 7: The training was underpinned by aspects of social cognitive theory and aimed to enhance teachers' self-efficacy in relation to PAL[16]. As such it drew from two prominent behaviour change techniques: barrier identification and modelling/demonstrating behaviour[17]. With the former, teachers were encouraged to identify barriers that might impact their ability to implement PAL and plan ways to overcome these. With the latter, the trainers demonstrated a plethora of PAL teaching strategies that teachers could employ in their lessons.

7) Line 217-221: too much repetition of 'the assistant head teacher'

This section has been edited to reduce repetition.

8) Line 353 – mouthguards and sport shoes purchased by pupils. It seems concerning that additional resources were being purchased by students. It might be worth elaborating on this as an issue in the Discussion.

Some text has been added to the pilot study reflection section to address this point: “Some students reported purchasing sports shoes and mouthguards for PAL; none of the strategies included in the PAL training involved students changing clothing/shoes or using mouthguards. It is conceivable that when completing the follow-up questionnaire some students considered P.E. lessons in their appraisal of PAL and reported sports shoes and mouthguards purchased for this.”

Reviewer 3

1) Page 7, Line 155/Page 8, Lines 156-157. There are concerns in many education circles about the measurement of students' weight at schools being problematic (alongside BMI being an inaccurate measure). Were there procedures used to minimise these concerns around students having their body weight measured? This could be a valuable inclusion.

We acknowledge these concerns and our measurement protocols take this into account. The research staff that carry out the height and weight measures receive regular training on how to take height and weight measures sensitively and accurately. At the school we set up the height and weight measurement stations in a separate area of the classroom (away from the group completing their questionnaires) and the stations were spread out. As such, students were not able to see any of their classmates' results. A weighing scale with a remote display was used – so the student being weighed was not able to see their weight measurement. We did not vocalize any of the student's height or weight measures – if a student asked what their weight measurement was, we showed them the written result on their data sheet.

We have added the following line of text to the feasibility study methods section: The measurement stations were set up so that results were not visible to anyone except the measurement staff.

2) Page 8, Lines 158-164. The rationale for using the triaxial wrist-worn accelerometers was based around 'energy expenditure' (& participation compliance), yet there was no data based specifically on energy expenditure in the results sections. Does this mean that the accelerometer intensity data presented is not valid? This needs to be looked into.

We used energy expenditure estimates to classify activity levels as sedentary, light, moderate and vigorous. The text in the manuscript which outlines this information reads as follows:

"ENMO thresholds were used to classify activity intensities: time spent at 0-30 ENMO was classified as sedentary activity (equivalent to 1-1.5 METs); 30-210 ENMO as light-intensity activity (1.5-4 METs); 210-500 ENMO as moderate-intensity activity (4-7 METs), and above 500 ENMO as vigorous-intensity activity."

3) Page 9, Line 194. Again, think classifying as overweight/obese based upon BMI is problematic. Perhaps a limitation.

We classified student weight status using BMI percentiles (not BMI). BMI percentiles are calculated using the height, weight, age and gender of a child and the result is expressed relative to other children of the same age and gender, using child growth charts. As such, the way we have classified weight status considers more than just height and weight. We acknowledge the limitations of BMI as a measure of overweight/obesity, but clarify that BMI was not an outcome of this study – we have used it only to describe the sample. As such, it does not weaken the findings or conclusions.

4) Page 9, Lines 197-198. Perhaps need to specify somewhere in the manuscript the implications of approximately 5 teachers not attending both sessions on the design/outcomes of the study. Similarly, later in the pilot study section (line 310).

We recognize that teachers not attending both training sessions could have implications for intervention outcomes. However, if the intervention were to be rolled out to other schools, we anticipate that in most instances a proportion of teachers at any one school would not be able to attend both training sessions (due to personal commitments or other responsibilities). As such, our findings have greater external validity and provide an accurate reflection of the outcomes we would observe in a real-world situation. The following paragraph has been added to the overall discussion: In both studies, a small number of teachers were unable to attend both training sessions which may have influenced intervention outcomes. It is realistic that ay any school receiving the intervention, a proportion of staff would be unable to attend both training sessions. As such the external validity of the findings is supported.

5) Page 9, Lines 197-202. Statistics were mentioned of improvement, it would be worth clarifying if these improvements were 'statistically significant'.

Due to the preliminary nature of the studies, we did not carry out a formal power calculation for sample size. As such, the study was not powered to detect significance and it would not be appropriate to perform tests of statistical significance on the data. Instead we have provided descriptive statistics.

6) Page 10, Lines 203-210. It appears as though there were a number of challenges for the teachers to deliver this type of intervention (see overall summary earlier in review). Many students in schools will also look for any opportunity to get out of 'English/Mathematics tasks/learning'. These training issues appear to be major limitations. It would be worthwhile trialling PAL in any non-movement subject areas in which the students genuinely want to participate and learn in.

We do not consider the training issues the reviewer refers to as limitations, but as indications of feasibility and acceptability of the intervention and PAL (i.e., valuable findings, given the aims of studies). Assessing feasibility and acceptability were two key aims of the studies, and the challenges the teachers reported are important findings which allow us to address those research aims.

The reviewer suggests trialing PAL in non-movement subject areas in which students really want to participate in, which we did in the pilot study. However, the number of lessons delivered for the subjects tends to be smaller than for maths and English and therefore excluding these subjects could have substantial implications for the dose of PAL received by the students.

7) Page 11, Line 238. It states 'teacher acceptability should be explored further', yet for the feasibility study to inform the pilot study, it would have been better from an educative perspective to focus on the training regime, preparation of both students and teachers.

Thank you for this suggestion. Teachers in the feasibility study reported a mixture of positive and negative responses regarding their experiences of delivering active lessons. As such, we feel our suggestion to further explore teacher acceptability in the pilot study was appropriate.

In addition, the feasibility study teachers responded positively to the training sessions – as such, prior to the pilot study, it didn't appear that an extensive revision of the training sessions was needed.

We feel the results from the pilot study have led us to the reviewer's suggestion that the content of the training sessions needs to be reviewed, which we will do prior to further implementation and evaluation.

8) Page 12, Lines 267-270. More detail into why it was not possible to blind is required. This is a major limitation.

We respectfully disagree that the lack of blinding of measurement staff represents a major limitation in these studies.

Not blinding measurement staff can cause problems when the measurement staff have the opportunity to influence the data. In these studies the students self-completed questionnaires (i.e., they were not coached by measurement staff on what answers to write) and measurement staff were only involved with fitting the students with the accelerometer (they did not encourage or discourage physical activity during the week that students were wearing the monitors). We acknowledge that the assessment of time on task could have been influenced by lack of blinding, but the purpose here was to assess time on task during a PAL, making it impossible to blind the assessor.

In regards to why it was not possible to blind measurement staff: participants at the intervention school completed some extra follow-up questionnaire measures, which were not relevant for control school participants (such as questions about taking part in active lessons). These additional measures required an introductory explanation from the research assistant (which measurement staff would have heard), and sometimes students had clarification questions about questionnaire items (which measurement

staff would have addressed). As such, it was not possible to blind the measurement staff to intervention and control school allocation. We believe that this has had limited impact on the findings from the two studies presented here, but acknowledge that a potential fully-powered trial would benefit from efforts to blind measurement staff.

In the interests of keeping the word count as low as possible, in the article we have mentioned differences in follow-up measures as the reason for not blinding measurement staff. We have also added the following comment to the strengths and limitations section of the overall discussion: Finally, we do not believe that lack of blinding of measurement staff has impacted the conclusions drawn from these studies, but acknowledge that a potential fully-powered trial would benefit from efforts to blind measurement staff.

9) Page 12, Paragraph 3. Please clarify if the schools were different in the pilot study to the feasibility study. This is important information to include.

The school that participated in the feasibility did not participate in the pilot study as well. The following text has been added to the recruitment section of the pilot study to clarify this: (the school that took part in the feasibility study was not invited to participate in the pilot study)

10) Methods overall. It states 'full-scale randomised controlled trial'. Yet in the methods there are only two schools. Please justify this amount of schools from the literature for a randomised controlled trial. Was there a power analysis conducted? This all needs to be included to justify the approach to the research. Is full-scale necessary?

In the introduction to the pilot study we state the following:

“we sought to extend our previous work and explore the potential value of conducting a full-scale randomised controlled trial.”

We apologise if this gave the impression that the pilot study is a full scale randomized controlled trial - in carrying out our pilot study we are exploring if there is value in carrying out a full-scale trial. The purpose of pilot testing is to run a smaller version of an anticipated full trial, which is what we did.

We mention in the feasibility study that we did not conduct a power analysis due to the preliminary testing focus of the study. We mention in the pilot study that we followed the same procedures as for the feasibility study.

11) Level of teaching experience of the teachers is important in this section (& at the school being measured). This could have a major bearing on how open minded the teachers were to the PAL intervention and not being entrenched in particular approaches (non PA).

We agree with the reviewer – the following text is included in the manuscript:

“Teacher feedback suggests that training acceptability is related to teachers’ experience delivering PAL. In the pilot study, teachers delivering PAL more regularly rated the intervention more poorly than less experienced teachers. A PAL intervention targeting teachers not regularly delivering PAL may be more acceptable.”

During preliminary data analysis, we explored relationships between total years of teaching experience and different indicators of intervention acceptability - the results did not indicate a clear/consistent association. As such, we did not include the information on this in the manuscript.

12) Page 15, Paragraph 2. From an educator perspective, it appears as though the students were taking advantage of the PAL intervention to escape learning.

We agree with the reviewer that this may have been happening. Within the manuscript we provide the following information related to the reviewer’s reflection:

- **Feasibility study results**: “Teacher-reported barriers included disruptive behaviour, lethargy and off-topic chatting, challenges re-focusing students after an active portion of class”
- **Feasibility results (from students)**: “Negative comments about PAL included lethargy (12%), more disruptive behaviour (9%), and less work achieved (12%; text box 1)”
- **Pilot study results**: “Students however also commented that during PAL some students messed around more and didn’t focus on work, and work was easier to do when sitting down”

13) Page 16. A really important methodological point to have considered is 'level of experience facilitating physical activity or physical education activities' etc. Not in terms of the PAL (specified on page 18, paragraph 1), yet in terms of their careers. Also, determining teachers' beliefs, experiences and intentions around physical activity prior to delivery. This could help plan for a lot of the barriers identified.

The training was adapted to teacher experience as much as possible, and more experienced teachers were encouraged to share their experiences with those who had limited experience. For example, at the start of the first training session, teachers were asked to identify barriers and solutions to PAL within a socio-ecological, school-based framework. This enabled teachers to identify where challenges may arise and then problem solve, as a group.

14) Page 18, Paragraph 1. The 'acceptability of the training was not demonstrated in the feasibility study'. Please correct.

We have reviewed paragraph 1 on page 18 - the sentence the reviewer highlights currently reads ‘While acceptability of the training was demonstrated in the feasibility study and is reported elsewhere[16, 37, 43], feedback from teachers in the pilot study was less positive.’

The wording of this sentence is as we intended. We apologise if we have misunderstood the reviewer’s comment.

15) Page 18. Important consideration for implementation is 'to include teachers in developing the study/design that would be anti-PE or who normally avoid facilitating physical activity where possible.' Gaining such teachers' thoughts would be invaluable to implement such a project successfully.

We appreciate this suggestion and we do not disagree. However, based on the results of this study, we cannot make this specific conclusion. We do not have data that tell us that teachers who are anti-PE were least engaged/did not implement the intervention. To clarify, we did work with a number of teachers (including those teaching non-PE subjects) in the design of the intervention and training programme to support successful delivery.

16) Page 20. Strength was definitely the formative work for interventions. Well done.

Thank you for your positive feedback.

17) Page 20. Perhaps some of my earlier points mentioned relating to limitations here. It still wasn't clear the relationship between rationale/reliability of the accelerometers (energy expenditure) and the results findings (various intensities).

Accelerometers measure movement intensity, which is related to energy expenditure (more intense movement = higher energy expenditure). The MET scale is a way of expressing the energy expenditure associated with different activities. 1 MET describes the energy expenditure of an average individual at rest. If that average individual engages in an activity that requires double the amount of energy expended at rest, that activity has an energy expenditure value of 2 METs. An activity requiring three times the energy expenditure at rest has an energy expenditure value of 3 METs, and so on. Sedentary activity corresponds to a MET value of 1, light intensity activity corresponds to a MET value of 1-3, moderate intensity activity corresponds to a MET value of 3-6, and vigorous intensity exercise

corresponds to a MET value of >6. These thresholds are well-established in the field of health and exercise science.

From the accelerometer data we are able to calculate energy expenditure and the associated MET value for each recorded data point. As such, we are able to sum the total time spent in sedentary, light, moderate and vigorous intensity activity.

18) Page 20, Conclusion, Line 462. Yes, but teacher acceptability was also 'not demonstrated'. This should be re-written.

In the feasibility study, teacher's acceptability of the training was demonstrated; in the pilot study, teacher's acceptability of delivering active lessons was demonstrated. The training sessions and the experience of delivering active lessons were two different aspects of the study, and we assessed feasibility and acceptability of both. Across the feasibility and pilot studies we demonstrated both – one aspect in each. As such, as well feel this statement is appropriate.

19) Page 20, Conclusion, Line 464. Was the intervention really acceptable to the students? I think there could be more scrutiny here around students looking to escape learning practices (what are the implications of being behind in the curriculum?). Were there gaps between communication between school leaders and teachers? E.g. School leaders wanting to shine by aligning with a major health initiative, teachers wanting to shine to the school leader by participating, yet in reality the teachers perhaps not wanting their classrooms distracted by the physical activity integration? Could this have led to there being some positives about the PAL integration? Some considerations around the findings, processes here.

We have divided the reviewer's comment down and addressed the statements:

- **Was the intervention really acceptable to the students?**
    - Based on quantitative and qualitative feedback, we have interpreted that students had a positive response to participating in physically active lessons. The students had the opportunity to provide positive and negative feedback on participating in active lessons, and our synthesis of the student responses indicated a primarily positive response.

- **I think there could be more scrutiny here around students looking to escape learning practices**
    - Observation data collected during active lessons did not suggest that students looked to escape learning practices.
    - We refer to student's behavior during active lessons in multiple places within the manuscript:
        - **Feasibility study results**: "Teacher-reported barriers included disruptive behaviour, lethargy and off-topic chatting, challenges re-focusing students after an active portion of class"
        - **Feasibility results (from students)**: "Negative comments about PAL included lethargy (12%), more disruptive behaviour (9%), and less work achieved (12%; text box 1)"
        - **Pilot study results**: "Students however also commented that during PAL some students messed around more and didn't focus on work, and work was easier to do when sitting down"

- **Were there gaps between communication between school leaders and teachers? E.g. School leaders wanting to shine by aligning with a major health initiative, teachers wanting to shine to the school leader by participating, yet in reality the teachers perhaps not wanting their classrooms distracted by the physical activity integration?**
    - Within each school, feedback from the assistant head teacher and teaching staff was consistent.

- **Could this have led to there being some positives about the PAL integration?**

- It's possible that teachers responded in a socially desirable way when reporting their active lesson efforts. The activity monitor data allowed us to see that activity hadn't increased during lessons identified by teachers as active lessons. A comment about teachers potentially responding in a socially desirable way has been added to the pilot study discussion.

20) Page 24, References. Could consider research on teachers (& early teachers e.g. PSTs) in how they report on difficult schooling experiences related to physical activity and physical education (PE) as pursuing non-PE specialist teaching. Early PE experiences can be tough for some to overcome. John Haynes from UNE wrote a piece on this and there are some snippets within the GET-PE study.

We agree that this is good literature to consider, however, a more extensive discussion around this specific barrier is beyond the scope of this article.

21) As it stands, the title appears as though the school recruitment in the UK is more widespread than in reality. Authors could consider perhaps reframing the title to: 'Piloting physically active lessons in UK secondary schools...' or something similar to note the relatively small sample.

Thank you for this suggestion. We appreciate the reviewer's suggestion of using the phrase 'piloting physically active lessons…', however, in accordance with trial registration guidelines on article titles, we prefer to mention the feasibility study and the nature of the pilot trial (cluster-randomised). We worked on different versions of the title, but the authors collectively agreed that the current title provides the best overview of the work conducted.

We hope the reviewer will consider that the inclusion of 'feasibility' and 'pilot trial' in the title, and the details of the numbers of participating schools in the abstract is enough information for readers to understand the preliminary nature of the work before reading the full article.

Reviewer 4

1) Line 27. Please insert 'was' after training.

This text has been added

2) Throughout manuscript please look into the use of apostrophe with students. e.g. Line 39 is students' rather than student's

Apostrophes have been reviewed and corrected as needed.

3) Would be good to insert a few brief examples of programme components in brackets somewhere to showcase brief elements of what the physical activities entailed (this could help generate interest from physical activity professionals with specific interests). An example would be line 37

We have provided examples in a supplementary file. We have also added some examples to the methods section of the abstract.

4) Line 193 and other incidences. Better to start sentence with N=99 than 99. Perhaps just refer to 92% on this line rather than another number.

Instances where a sentence started with a digit have been adjusted to start with a word.

5) Line 194. Would be good to specify the standard deviation here for 13.0 years.

The standard deviation has been added.

6) Page 10, Line 211. Perhaps just percentage here.

The text has been adjusted.

7) Line 306. Again, go with N= to start sentence and referring to 92% here will be fine without the brackets and referring to 205.

Instances where a sentence started with a digit have been adjusted to start with a word.

P2 Line 34 and line 38: Need an adjective before acceptability e.g. poor, excellent etc.

Adjectives have been added.

P4 Line 60 and line 63: Need space before citation brackets. Other occurrences in the manuscript need amending.

We have made sure the spacing between citations and text is consistent throughout the manuscript.

P4 Line 62: Insert comma after ‘interventions’. And remove comma after ‘hours’.

This has been corrected

P4 Line 70: Should this instead read something like ‘perceived to be the most-feasible’, acceptable etc.? As to my knowledge this is the first PAL intervention in secondary settings, in the UK at least.

The text has been edited.

P5 Line 90: It would be good to set out more clearly in this paragraph exactly what each study intended to do, rather than blending the aims. This is important because as you read on the purpose of the feasibility study isn’t overt for the reader.

A sentence has been added to clarify the differences between the feasibility and pilot studies.

P6 Line 111: Why 120 students? Was this a pragmatic choice due to resource/manpower available? If so it’s worth stating this. Also why Year 7 and Year 9 students?

The reason for using 120 students is mentioned in the feasibility study’s ‘recruitment’ paragraph: “The study’s feasibility focus meant that a formal power calculation was not necessary to inform sample size; a sample of 60 participants per year is consistent with samples of similar studies[15].”

A line of text has been added to the first paragraph of the recruitment section to indicate that the head teacher chose years 7 and 9 to participate in evaluation measures.

P6 Line 120: Assume this is five PAL lessons? Suggest inserting ‘PAL’ to clarify.

We did not intend to refer to PAL lessons. The information the reviewer refers to is provided so that readers understand how often students receive maths and English classes and therefore the potential exposure of students to active lessons.

P6 Line 123: Where was the training delivered? What time in the school year? Did teachers have to be bought out of usual teaching time? A few of these details would be useful.

This information has been added to the 'intervention' section of the feasibility study methods.

P8 Line 177: I'm a little confused as to why time-on-task was included as a pre-post measure. Usually it is used to measure TOT during or post PAL. Did you expect TOT to change due to children having taken part in PAL? Also what was the protocol for the observation, did you for example focus on one table at a time and work around in a systematic order focusing on one student at a time?

Information has been added to the 'time on task' section of the feasibility study and pilot study methods sections to clarify the observation protocol. In the pilot study methods section, we have clarified that the research team asked to observe active lessons at follow-up (to allow comparison of TOT during desk-based lessons at baseline and during more active lessons at follow-up).

P11 Line 231: Effectiveness could be re-worded given that the study was not powered to detect change in these variables and due to a lack of control group (as you acknowledge).

We have made sure all mentions of 'effectiveness' are preceded by the word 'preliminary'.

Page 12 Line 262: Why did you decide to recruit three schools and were they selected based on size, SES, ethnicity etc.? It would be helpful if some information on the schools were presented, for example what was their Ofsted rating, school size, SES etc.

Information on the number of school selected has been added to the pilot study methods 'schools' section: "We aimed to recruit three schools - two intervention (to test whole-school delivery of the intervention in different settings) and one control (to test the acceptability of school-level randomisation)."

The school inclusion criteria are specified in this section as well: 26 non fee-paying, mixed gender, secondary schools in the East of England were emailed study information and invited to participate

Additional school-level statistics have been added to Supplementary Table 2.

P16 Line 359: I think you should mention the issue of follow-up students being different to baseline students for some of the measures. Was this due to practical reasons? An interesting point re. feasibility of data collection in this population perhaps.

Apologies if we have not interpreted this comment correctly:

We experienced a small amount of participant drop-out, although in all instances, over 90% of participants were retained between baseline and follow-up so differences between the baseline and follow-up sample are likely to be minimal.

When statistics are presented for baseline and follow-up values, they reflect data from participants with complete data at both time points. In table 2 the different N's for each measure indicate different numbers of participants providing sufficient or complete data for each measure, but the summary statistics within the table are based on individuals with data at both time points.

P18 Line 392: Did the teacher's comment on the duration of the training and if they required a top-up, or if they felt competent enough to deliver PAL once trained? My experience in primary teachers is that even a full day of training is insufficient. I wonder therefore if 4 hours of training is sufficient to change teacher's pedagogy? It's quite a big step change for them, unlike for example integrating a few standing desks in the corner of the room.

In the overall discussion we comment on the possibility that the teachers may have still been mastering PAL strategies in the feasibility study, although the teachers did not indicate this was an issue in their feedback. We have also added some information regarding the length of the training sessions to the 'preliminary effectiveness' section of the overall discussion.

Table 1. Were the follow-up measures taken whilst the teacher's were still implementing PAL? As from table 3 it appears you have accelerometer data from PAL lessons, but I can't see where in the main text of the manuscript that you describe how you collected this data. This also relates to your point in the discussion, where I think you are right in stating that it might be a stretch for accelerometers to capture short periods of PAL lessons among the noise of an overall day. Therefore one limitation is that you weren't able to collect direct observation of what they were doing in a sample of the lessons, which might help explain the small dose of PA received.

In the pilot study methods we mention: "During the student accelerometer assessment at follow-up, teachers were given their personalised timetable and asked 'please circle which of the listed Year 7 and/or 9 classes were (or will be) delivered as an active lesson.'" As such, for the week when students were wearing accelerometers, we were able to work out when they were in active lessons. This allowed us to extract the portions of accelerometer data that corresponded to the times the students were in PAL, and calculate average activity during PAL.

We have added text to the overall discussion regarding the potential benefits of further direct observations: . Across both studies, teachers were advised that any non-seated activity was considered an 'active lesson' - as such, the intervention may be too dilute for measurable impact using wrist-worn accelerometers; classroom observations of PAL (beyond assessing time on task) may have aided our interpretation of the findings.

VERSION 2 – REVIEW

REVIEWER	Nicole Nathan The University of Newcastle NSW Australia
REVIEW RETURNED	20-Nov-2018

GENERAL COMMENTS	The authors have made a great effort in addressing reviewers comments.

REVIEWER	Emma Norris Centre for Behaviour Change, UCL
REVIEW RETURNED	Centre for Behaviour Change, UCL 16-Nov-2018

GENERAL COMMENTS	Thank you for addressing my comments - happy to accept. This is a valuable contribution to the field which is currently lacking research in secondary schools.

REVIEWER	Doctor Brendon Hyndman Charles Sturt University, Australia
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	The authorship team have satisfactorily responded to and addressed the reviewer queries with the manuscript. Before publication, it would be valuable for the authors to provide more clarity on 'feasibility/acceptability' within the manuscript. It should be acknowledged that the core business of education/schools is for 'students to learn'. Therefore, if teachers indicate that students are not learning as much during PAL, students are lethargic and resistant-- is PAL really acceptable and feasible to schools and educators? The feasibility/acceptability is very much focused on a 'health perspective' with the intervention delivery, which is the obvious lens/angle of the research. Yet further clarity on these aspects would help an educational audience with the conclusions generated.

REVIEWER	Ash Routen Loughborough University, UK
REVIEW RETURNED	16-Nov-2018

GENERAL COMMENTS	Good job addressing my remarks. Other than the below I have no further comment: Line 29: Should read ‘including’ not ‘included’.

VERSION 2 – AUTHOR RESPONSE

We thank the reviewers for their time in re-reviewing our article. Below are our responses to the comments raised from this second review:

Reviewer 4: Line 29: Should read ‘including’ not ‘included’.

Our Response: Line 29 is part of the abstract (the 'intervention' section), which currently reads: "Teachers were made aware of how to integrate movement into lessons; strategies included students collecting data from the environment for class activities, and completing activities posted on classroom walls, instead of sitting at desks."

Reviewer 4 has suggested changing ‘included’ to ‘including’. On re-reading this sentence ‘included’ appears to be grammatically correct.

Reviewer 3: Before publication, it would be valuable for the authors to provide more clarity on 'feasibility/acceptability' within the manuscript. It should be acknowledged that the core business of education/schools is for 'students to learn'. Therefore, if teachers indicate that students are not learning as much during PAL, students are lethargic and resistant-- is PAL really acceptable and feasible to schools and educators? The feasibility/acceptability is very much focused on a 'health perspective' with the intervention delivery, which is the obvious lens/angle of the research. Yet further clarity on these aspects would help an educational audience with the conclusions generated.

Our response: The following text has been added to the overall discussion ('feasiblity/acceptability of PAL training' section): "Teacher’s concerns regarding the lack of learning associated with PAL strategies must be an important consideration in the design of future PAL interventions. Student learning is the core focus of schools and implementation of PAL is likely to be contingent on teachers perceiving that PAL supports this goal."