

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)	Do school-based smoking preventive interventions have unintended effects? Post hoc analysis of the Focus cluster randomized controlled trial
AUTHORS	Kjeld, Simone G.; Thygesen, Lau; Danielsen, Dina; Jensen, Marie; Krølner, Rikke; Pisinger, Charlotta; Andersen, Susan

VERSION 1 – REVIEW

REVIEWER	Lim, Kuang Hock Institute for Medical Research
REVIEW RETURNED	24-Nov-2022

GENERAL COMMENTS	The manuscript is of public health interest which investigated on the unintended outcome of smoking intervention, however, the authors should address several comments as below:- Abstract The introduction in this section is too long, kindly shorten it. Introduction The author should include more literature related to the outcome of similar studies to enrich and justify the need of current write-up. The write-up should need to re-arrange (en. Paragraph 2 should be in paragraph 1) to ensure to flow of ideas in the text. Methodology How the sample size, randomization and measure of data collection should be described in more detail in the text. The hypothesis should not be too long in the manuscript. The fit of multilevel logistic regression should be tested to ensure the results are valid. Discussion More in-depth discussion with similar previous studies with enrich the content of the manuscript.
-------------------------	--

REVIEWER	Crone, Mathilde Leiden University Medical Center
REVIEW RETURNED	11-Dec-2022

GENERAL COMMENTS	It is an interesting paper in which the researchers looked at the potential negative effects of a smoking prevention programme. No
-------------------------	--

	real adverse effect were found in the total group. Although there were some remarkable differences regarding boredom and stress. With the intervention group reporting somewhat less boredom and stress and in subgroups seeing significant effects on boredom and stress. In the methods section the authors describe that the programme that was developed was intended to also accommodate the potential adverse effects of the programme. With several activities focussed at decreasing boredom and stress. It would be good to also elaborate on this in the introduction. Although there are no statistically significant effects, The authors discuss some tendencies to significance, such as the effects on cannabis use and boredom and stress. however as it consists partly of imputed data the outcomes are already an approximation. I recommend therefore to be careful with using the word significant tendency. In line with this it would be good to say more about the missing data. How large was the part that was missing? The outcomes of the official trial are not get published, I was wondering whether the authors should not distinguish between the smokers that quitted smoking and the ones who didn't. As it could be hypothesised that boredom and stress is essentially a problem for the ones that kept smoking and were not able to smoke at school because of policies. While on the other hand the use of alternatives may be more for the ones that quitted. But it also depends on how many quit or changed in their smoking behaviour. Why only including the daily smokers in the analyses and not the occasional smokers? As analyses are conducted on only the daily smokers should differences at baseline not be assessed for only the group of smokers? How many schools had a fully implemented programme and how many students did it concern? And what did the schools implement that partly implemented the programme? The authors discuss the stigmatisation finding or the tendency of the intervention group feeling more stigmatised. And I agree that more research should be conducted. The question is however not only whether it is an adverse effect , as there are also discussions that stigmatisation or denormalisation can lead to more carnation. Secondly they discuss the adverse effects they did not look at but might have occurred. They state that equity harm is however not likely as the study was conducted In a lower ses group of adolescents. However there are examples that even - if a study is done in a lower ses – setting the one with a higher socioeconomic position in this setting profit more (https://pubmed.ncbi.nlm.nih.gov/22350193/).
REVIEWER	Tessier, Katelyn M University of Minnesota No competing interests. I will disclose that I am a statistician on several tobacco related clinical trials in the U.S.
REVIEW RETURNED	06-Feb-2023

GENERAL COMMENTS	The study and design are very interesting and valuable to tobacco cessation research. This manuscript is well-written and lays out the design and methods pretty clearly. However, there are some areas that need further clarification that will improve this manuscript. Abstract: -The abstract states 1,028 students were randomized to the intervention, but in the main text and results it states 844. Please update. Introduction: - It seems uncommon that this paper would be published before a paper on the results of the primary and secondary outcomes of the RCT (stated in lines 27-37 on page 4 of the manuscript). Please provide rationale or further clarification. Methods: -Within the section 'Study population, recruitment, and randomization' starting on page 5 of manuscript at line 49, there is no lower limit of age listed. Although it is listed in the study protocol published article, please revise to include the targeted age of 16-25 years. -Within this same section, there is no description of number of schools randomized to each group (of the 14 schools). Please add. Additionally, why were 8 schools randomized to intervention and 6 to control instead of equal allocation? -In the Statistical Analysis section, what variables were used for multiple imputation? Please be explicit in this section. - In this same section, please update to explicitly state what levels were included in the multilevel logistic regression models. The way it is currently written, the levels are listed within the per protocol portion of the analysis, but I am assuming that levels for school, class and individual were used for the ITT analysis as well. -Were generalized linear mixed models considered to include information from both follow-up time points? If not, please provide rationale for why only the 5-month follow-up time was included. Additionally, in the Statistical Analysis section, please be more specific on what time points were used. -The Statistical Analysis section states that ITT approach was used for main analyses including all students with responses at baseline but then later it states that the analyses were performed for students who were daily smokers at baseline. Please clarify, and consider whether there is any necessity for presenting results on all students and not just daily smokers. Results: -Please include p-values in Table 1. If all were non-significant, it is fine to leave p-values off but consider including a footnote below Table 1 explaining this. -Please include a statement or footnote for why the discrimination subscale wasn't formally analyzed in Table 2. I assume it's due to the small percentage, but this should be explicitly stated somewhere. Similarly, please be explicit in Table 4 as well (for regular cannabis use and felt stigma).
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Kuang Hock Lim, Institute for Medical Research Comments to the Author:

The manuscript is of public health interest which investigated on the unintended outcome of smoking intervention, however, the authors should address several comments as below:-

Abstract

The introduction in this section is too long, kindly shorten it.

Response: *We have sought to shorten the introduction and rearranged it as you have suggested.*

Introduction

The author should include more literature related to the outcome of similar studies to enrich and justify the need of current write-up.

Response: *We have re-arranged the introduction and described the previous literature more in-depth. However, there is generally very limited literature concerned with unintended effects of public health/smoking interventions in this setting.*

The write-up should need to re-arrange (en. Paragraph 2 should be in paragraph 1) to ensure to flow of ideas in the text.

Response: *We have re-arranged the introduction and hope that the flow is now easier to follow.*

Methodology

How the sample size, randomization and measure of data collection should be described in more detail in the text.

Response: *Thank you for this comment, which we agree with. We have updated the section with more information: "Schools were recruited in two rounds over the course of three months (January 2018 to April 2018 and again from January 2019 to April 2019) to ensure that the schools had sufficient time to implement the intervention before school start in August. Seven VET schools were randomly assigned to the intervention group (n=4) or control group (n=3) during the first round. Because it showed difficult to recruit VET schools with 'technology, construction, and transportation' courses, the second round included PBE schools. In this round, three VET schools and four PBE schools were randomly assigned to the intervention group (n=4; two PBE schools and two VET schools) or control group (n=3; two PBE schools and one VET school). Subsequently, 14 Danish schools were enrolled in this study (N(students)=1,659), of which 844 students were assigned to the intervention group (response proportion 76%) and 815 to the control group (response proportion 75%). In total, 432 students in the intervention group and 374 students in the control group were lost to follow-up at five months after baseline. A more detailed description of the sample flow is reported elsewhere.¹³"*

The hypothesis should not be too long in the manuscript.

Response: *We have now amended the introduction.*

The fit of multilevel logistic regression should be tested to ensure the results are valid.

Response: *Thank you for this comment. We have not included these fit tests of the model. Hence, most tests for model fit would indicate that the model does not adequately fit the data, primarily because the model does not include all variables that explain the variation in the data. This is especially true in larger data sets.*

Discussion

More in-depth discussion with similar previous studies with enrich the content of the manuscript.

Response: *Thank you for this comment, we have included a more thorough discussion with previous literature to the manuscript.*

Reviewer: 2

Dr. Mathilde Crone, Leiden University Medical Center Comments to the Author:

.It is an interesting paper in which the researchers looked at the potential negative effects of a smoking prevention programme. No real adverse effect were found in the total group. Although there were some remarkable differences regarding boredom and stress. With the intervention group reporting somewhat less boredom and stress and in subgroups seeing significant effects on boredom and stress.

In the methods section the authors describe that the programme that was developed was intended to also accommodate the potential adverse effects of the programme. With several activities focussed at decreasing boredom and stress. It would be good to also elaborate on this in the introduction.

Response: *Thank you for this comment, we have sought to elaborate more about the mechanisms related to boredom and stress in the introduction (which has been rewritten and restructured), while we address in more detail how the intervention sought to accommodate this in the Methods section.*

Although there are no statistically significant effects, The authors discuss some tendencies to significance, such as the effects on cannabis use and boredom and stress. however as it consists partly of imputed data the outcomes are already an approximation. I recommend therefore to be careful with using the word significant tendency. In line with this it would be good to say more about the missing data. How large was the part that was missing?

Response: *We have rephrased in places where we indicate statistical significance. However, it is not the purpose to state that there are tendencies towards significance, merely that there are some tendencies in the data – although it should be noted that these tendencies were non-significant. We have moreover included more information in the Methods section as you have suggested: “In total, 432 students in the intervention group and 374 students in the control group were lost to follow-up at five months after baseline.”*

The outcomes of the official trial are not yet published, I was wondering whether the authors should not distinguish between the smokers that quit smoking and the ones who didn't. As it could be hypothesised that boredom and stress is essentially a problem for the ones that kept smoking and were not able to smoke at school because of policies. While on the other hand the use of alternatives may be more for the ones that quit. But it also depends on how many quit or changed in their smoking behaviour.

Response: *Thank you for this reflection. We have published the outcomes of the trial now and referenced it in the study. We agree that these mechanisms may in fact be present, however, as is shown in the effect study, not many students changed their behavior, and we would not be able to distinguish between current smokers and those who quit smoking from baseline to follow-up due to lack of power.*

Why only including the daily smokers in the analyses and not the occasional smokers?

Response: *We decided to only include daily smokers, as we hypothesized that this was the group that would especially be affected by the structural smoking policies, whereas students who smoke occasionally would not have a significant part of their life at school restricted. Hence, we do not anticipate the occasional smokers to experience negative consequences similar to that of daily smokers. We also added the following sentence to the manuscript: “[...] we do not anticipate the occasional smokers to experience negative consequences similar to that of daily smokers. This could be attributed to their lack of self-identification as ‘smokers’ – but instead as individuals who smoke for a period,^{32 33} and therefore, they would not necessarily identify themselves as primary targets of the intervention.”*

As analyses are conducted on only the daily smokers should differences at baseline not be assessed for only the group of smokers?

Response: *Thank you for this comment. We agree that this would make more sense, and we have therefore decided to modify the baseline table to only include students who smoked at baseline.*

How many schools had a fully implemented programme and how many students did it concern? And what did the schools implement that partly implemented the programme?

Response: Thank you for bringing our attention to this. We have added a paragraph in the Methods about the implementation measure and also provided the number of schools and students within each category: "Implementation measure was assessed according to school principals' answers to a questionnaire about the delivered intervention components. According to their responses, the implementation measure was categorized into 1) implementation of the school- and class-level intervention components (intervention delivered in full; n=4 schools, 134 students), 2) implementation of some class-level components or withdrawal of components during the evaluation period (partial intervention; n=4 schools, 134 students), and 3) control schools (n=6 schools, 227 students)."

The authors discuss the stigmatisation finding or the tendency of the intervention group feeling more stigmatised. And I agree that more research should be conducted. The question is however not only whether it is an adverse effect, as there are also discussions that stigmatisation or denormalisation can lead to more carnation.

Response: Thank you for this important point. We have elaborated more about the stigmatization tendency, and the ongoing discussions about stigmatization in prior research and literature: "The potential stigmatization of students is an area of concern that warrants attention. Smoking-related stigma has been widely discussed in prior literature,^{29 30} with studies suggesting that initiatives aimed to increase smoking-related stigma could potentially have a positive impact on reducing the overall smoking prevalence or at least the desire to quit smoking.^{29 30 47} One hypothesized mechanism is that public stigma of smoking may lead to an internalized smoking-related stigma and hence, an increased desire to quit smoking.²⁹ Nevertheless, ethical considerations should be carefully considered within this context, particularly in light of the fact that some individuals may employ smoking as a coping mechanism for various challenges in their lives, such as stress reduction or alleviation of psychological symptoms.^{25 26} Research has also shown that self-stigmatization may negatively affect the mental health of individuals and help-seeking behaviors in the context of smoking.^{28 29} Therefore, efforts aimed at increasing perceived stigma among individuals who smoke could prove to be counterproductive."

Secondly they discuss the adverse effects they did not look at but might have occurred. They state that equity harm is however not likely as the study was conducted in a lower SES group of adolescents. However there are examples that even - if a study is done in a lower SES - setting the one with a higher socioeconomic position in this setting profit more (<https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fpubmed.ncbi.nlm.nih.gov%2F2350193%2F&data=05%7C01%7Csimk%40sdu.dk%7CB73a6d9b568e47f8888c08db342788cf%7C9a97c27db83e4694b35354bdf18ab5b%7C0%7C0%7C638161116625074255%7CUnknown%7CTWFpbGZsb3d8eyJWljiMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6Ikl1haWwiLCJXVCi6Mn0%3D%7C3000%7C%7C%7C&sdata=IJOrehwzdMxg33aQF9Tofz8cYj3X0UekYkYiS8g6xel%3D&reserved=0>).

Response: Thank you for this reflection, we have added this to the discussion: "It is worth noting that - even within the context of this setting with lower SES students - the intervention may have benefitted those with the highest SES to a larger extent, as demonstrated in previous research.⁴⁹ This aspect was beyond the scope of this particular study on unintended effects experienced by students who smoke daily, but is an important area for future research."

It should also be noted that the measure of SES available may not be a good enough reflection of the students' actual SES, as we only have information on parental SES. This is a good measure of SES in younger adolescents (such as those in primary school), however, many within this age group are above 18 years old, have their own income, and live away from their parents, and a proxy of their individual SES might be more appropriate.

Reviewer: 3

Dr. Katelyn M Tessier, University of Minnesota Comments to the Author:

The study and design are very interesting and valuable to tobacco cessation research. This manuscript is well-written and lays out the design and methods pretty clearly. However, there are some areas that need further clarification that will improve this manuscript.

Abstract:

-The abstract states 1,028 students were randomized to the intervention, but in the main text and results it states 844. Please update.

Response: *We apologize for this mistake and have changed the number to 844 in the abstract.*

Introduction:

- It seems uncommon that this paper would be published before a paper on the results of the primary and secondary outcomes of the RCT (stated in lines 27-37 on page 4 of the manuscript). Please provide rationale or further clarification.

Response: *Thank you for this comment, we completely agree. The paper concerning results of the primary and secondary outcomes has now been published, and we have updated the manuscript with a reference.*

Methods:

-Within the section 'Study population, recruitment, and randomization' starting on page 5 of manuscript at line 49, there is no lower limit of age listed. Although it is listed in the study protocol published article, please revise to include the targeted age of 16-25 years.

Response: *Thank you for this comment, we have added the following sentence: "Generally, students within these educational settings are between 16-25 years old, and the intervention was targeted this age group."*

-Within this same section, there is no description of number of schools randomized to each group (of the 14 schools). Please add. Additionally, why were 8 schools randomized to intervention and 6 to control instead of equal allocation?

Response: *Thank you for bringing our attention to this, we have added more information about the randomization process: "Schools were recruited in two rounds over the course of three months (January 2018 to April 2018 and again from January 2019 to April 2019) to ensure that the schools had sufficient time to implement the intervention before school start in August. Seven VET schools were randomly assigned to the intervention group (n=4) or control group (n=3) during the first round. Because it showed difficult to recruit VET schools with 'technology, construction, and transportation' courses, the second round included PBE schools. In this round, three VET schools and four PBE schools were randomly assigned to the intervention group (n=4; two PBE schools and two VET schools) or control group (n=3; two PBE schools and one VET school). Subsequently, 14 Danish schools were enrolled in this study (N(students)=1,659), of which 844 students were assigned to the intervention group (response proportion 76%) and 815 to the control group (response proportion 75%). In total, 432 students in the intervention group and 374 students in the control group were lost to follow-up at five months after baseline."*

To clarify, there are more intervention schools than control schools because we performed stratified randomization and not block randomization. The stratified randomization procedure was conducted to prevent imbalance between groups regarding type of school. However, within strata (the school types) had varying size (i.e., unequal number of schools), and it was by chance that more schools ended up in the intervention group.

-In the Statistical Analysis section, what variables were used for multiple imputation? Please be explicit in this section.

Response: *we added more information about the included variables for the multiple imputation model: "The main analyses were based on multiple imputations (MI) of missing data using several variables, which included school variables, sociodemographic variables, individual smoking-related behavior and beliefs about smoking, and smoking within social relations."*

- In this same section, please update to explicitly state what levels were included in the multilevel logistic regression models. The way it is currently written, the levels are listed within the per protocol

portion of the analysis, but I am assuming that levels for school, class and individual were used for the ITT analysis as well.

Response: Thank you for bringing our attention to this. We have included more information about the multilevel logistic regression models: “Baseline age, gender, and SES were included as covariates as well as the baseline value of the outcome if measured (regular use of smokeless tobacco, hookah, and e-cigarettes were measured at baseline) and the stratification variable, school type. Each model also included the class variable as a random effect to account for the clustering effect.”

Although we used the school level to the random allocation, we chose to adjust for clustering at the class level. This is because the intervention components were delivered at class level, and therefore, adjustment for class will take into account the highest clustering effect.

We have removed the information about how the implementation measure was retrieved to avoid confusion and instead added a description of this measure in the Measures section.

-Were generalized linear mixed models considered to include information from both follow-up time points? If not, please provide rationale for why only the 5-month follow-up time was included. Additionally, in the Statistical Analysis section, please be more specific on what time points were used.

Response: Thank you for this comment. We chose to not include information from the half-way follow-up. This data consisted of a questionnaire that was sent to students through SMS (text message) that resulted in a very low response rate, and we did not see fit to include this data in the paper. We included a sentence in the Statistical Analysis section stating that we used the five-months follow-up.

-The Statistical Analysis section states that ITT approach was used for main analyses including all students with responses at baseline but then later it states that the analyses were performed for students who were daily smokers at baseline. Please clarify, and consider whether there is any necessity for presenting results on all students and not just daily smokers.

Response: Thank you for this comment, we have rephrased the Statistical Analysis section and made clear that we solely included daily smokers at baseline.

Results:

-Please include p-values in Table 1. If all were non-significant, it is fine to leave p-values off but consider including a footnote below Table 1 explaining this.

Response: We appreciate your suggestion regarding the inclusion of p-values in Table 1. Given that our study is a randomized controlled trial, we opted not to include p-values in the baseline table. In a standard randomized study design, any observed differences would be regarded as chance occurrences (and therefore not something that should be adjusted for). Therefore, we did not see fit to present p-values in the baseline table.

-Please include a statement or footnote for why the discrimination subscale wasn't formally analyzed in Table 2. I assume it's due to the small percentage, but this should be explicitly stated somewhere. Similarly, please be explicit in Table 4 as well (for regular cannabis use and felt stigma).

Response: Thank you for this good suggestion. We have updated the tables with a footnote stating why these variables were not analyzed.

VERSION 2 – REVIEW

REVIEWER	Crone, Mathilde Leiden University Medical Center
----------	---

REVIEW RETURNED	03-Jul-2023
GENERAL COMMENTS	The authors have responded well to my comments and good to now also be able to read the original paper on the primary outcomes, which showed no overall effect and small effects on smoking behaviour among girls and schools that implemented the intervention fully. It is interesting to see that the level of implementation makes a difference. In this study on potentially unintended effects the authors found that in schools with a fully implemented intervention the students were more likely to report "felt stigma". They now provide a better discussion about this finding. It might be good to also link it to the outcomes of the main article, in which these school with fully implemented intervention had more quitters. A potentially interesting publication is also the one from Etter & Bouvier and from Hanewinkel on the smokefree class competition: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566022/ https://www.jstor.org/stable/20721171 Interesting, also for the discussion, is that the partial intervention schools are more likely to show lower stress levels (and also somewhat lower "felt stigma"). Do you have any information on what these schools implemented (that might explain these findings) as now the methods only report that they implemented some class-level components or withdrawal of components during the evaluation period? Furthermore I understand from their answers and methods that the school principal completed the implementation questionnaire . This might be sufficient for the school level activities but might not have sufficient detailed information on what has actually been done in the classes. Furthermore they adjusted for class level. I was however wondering whether then the number of daily smokers per class is sufficient for conducting this multilevel-analyses. And how many classes were then included?
REVIEWER	Tessier, Katelyn M University of Minnesota No competing interests. I will disclose that I am a statistician on several tobacco related clinical trials in the U.S.
REVIEW RETURNED	13-Jun-2023
GENERAL COMMENTS	My concerns/comments were addressed in this revision. However, I have one new concern/question for the authors. I apologize for this oversight in my first review. Results: -Please clarify why Table 2 has p-values but Tables 3 and 4 do not. The manuscript would be stronger with consistency of reporting results.

VERSION 2 – AUTHOR RESPONSE

Reviewer #3:

My concerns/comments were addressed in this revision. However, I have one new concern/question for the authors. I apologize for this oversight in my first review.

Results: Please clarify why Table 2 has p-values but Tables 3 and 4 do not. The manuscript would be stronger with consistency of reporting results.

Reply: We agree with the reviewer and have deleted the p-values from Table 2.

Reviewer #2:

The authors have responded well to my comments and good to now also be able to read the original paper on the primary outcomes, which showed no overall effect and small effects on smoking behaviour among girls and schools that implemented the intervention fully. It is interesting to see that the level of implementation makes a difference. In this study on potentially unintended effects the authors found that in schools with a fully implemented intervention the students were more likely to report "felt stigma". They now provide a better discussion about this finding. It might be good to also link it to the outcomes of the main article, in which these school with fully implemented intervention had more quitters. A potentially interesting publication is also the one from Etter & Bouvier and from Hanewinkel on the smokefree class competition: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566022/> <https://www.jstor.org/stable/20721171>

Interesting, also for the discussion, is that the partial intervention schools are more likely to show lower stress levels (and also somewhat lower "felt stigma"). Do you have any information on what these schools implemented (that might explain these findings) as now the methods only report that they implemented some class-level components or withdrawal of components during the evaluation period?

Reply: Thank you for these valuable comments and suggestions.

We know that all schools implemented the staff course, the edutainment session, the competition, and information about smoking cessation support while the tobacco policy was not implemented as intended at the partial intervention schools. We have amended the Discussion about stress level in partial intervention schools as follows (page 15):

"An explanation may be that the school tobacco policy was to a lesser extent implemented or not implemented at all, while the schools still received the staff course, the edutainment session, the competition, and gave information about smoking cessation support. Hence, students may still have been able to smoke during school hours (and subsequently, use smoking to e.g., cope with negative emotions), while also receiving sources of support and activities aimed at promoting social relationships."

Another explanation could be that the partial intervention schools had lower stress levels among their students at baseline. This information is lacking and is added to the Limitation section (page 18), as follows:

"Moreover, some of the measurements of the unintended effects were only measured in the follow-survey and not at baseline which precluded adjustment for potential baseline differences."

As suggested from the Reviewer, we have linked the findings regarding stigma to the outcomes of the main analysis (page 16):

"This is supported by the main analysis of Focus suggesting that full intervention schools have fewer daily and regular smokers at follow-up, while no effect was seen in schools that have only partially implemented the intervention."¹⁷

Further, in the section of Implications for practice (page 18-19), we have included the following:

"Prior analyses of Focus indicate that the school tobacco policy drives the reduction of smoking in the participating schools while the other components may support the implementation but cannot stand alone. For example, the students may experience the class competition as enjoyable⁵¹ which may

motivate them to adhere to the school tobacco policy but cannot stand alone. Studies among younger adolescents that have examined smoke-free class competition programs shows no long-term effects on smoking outcomes⁵² nevertheless it has not been investigated as add-on to implementing a strict school tobacco policy.”

Furthermore I understand from their answers and methods that the school principal completed the implementation questionnaire . This might be sufficient for the school level activities but might not have sufficient detailed information on what has actually been done in the classes.

Reply: We only included principal questionnaire data for the implementation measurement in the present study to form a simple school-level indicator of implementation. Since the submission of this manuscript, we have published a paper describing the development of comprehensive measurements of implementation underpinned by a conceptual framework based on principal, teacher, and student implementation questionnaires: Jensn, M.P., Krølner, R.F., Jørgensen, M.B. *et al.* Assessment of Delivery and Receipt of a Complex School-Based Smoking Intervention: A Systematic Quantitative Process Evaluation. *Glob Implement Res Appl* **3**, 129–146 (2023). <https://doi.org/10.1007/s43477-023-00084-5>

These measurements are currently being used to test the association to smoking outcomes. However, the principal data included in the present paper may represent the best source of information for a simple school-level indicator, since we struggled to invite and/or get responses from the teachers who were involved in the delivery of the intervention components such as the curriculum.

Furthermore they adjusted for class level. I was however wondering whether then the number of daily smokers per class is sufficient for conducting this multilevel-analyses. And how many classes were then included?

Reply: Multilevel models were utilized to take account of the clustering within classes. The number of classes included was 86 (this information is given in the flow chart of the main article). We did not include information on number of smokers per class but assumed instead that each student within a given class may have been influenced by the smoking habits in the class.

VERSION 3 – REVIEW

REVIEWER	Crone, Mathilde Leiden University Medical Center
REVIEW RETURNED	20-Nov-2023
GENERAL COMMENTS	The authors answered my questions regarding implementation and multilevel analyses. I have no further comments.