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)	What medical students think about measurement of their wellbeing: cross sectional survey and qualitative interviews
AUTHORS	Simons, Gemma; Effah, Raymond; Baldwin, David

VERSION 1 – REVIEW

REVIEWER	Erschens, Rebecca University Medical Hospital Tuebingen, Internal Medicine, Department of Psychosomatic Medicine and Psychotherapy
REVIEW RETURNED	05-Oct-2021

GENERAL COMMENTS	Introduction Thank you for the opportunity to review the manuscript 'What medical students think about measurement of their wellbeing and implications for pastoral support: cross sectional survey and qualitative interviews'. This manuscript investigates medical student`s opinion on the measurement of wellbeing regarding the frequency of availability for measurement, the format, type and purpose of measurement and with whom wellbeing should be discussed. Medical students from different years of Southampton University took part in an online survey and semi-structured interviews. Overall your manuscript is well structured and easy to follow. Your ideas regarding improving student well-being are well elaborated. Unfortunately, the introduction does not adequately introduce the rationale for your choice of research questions. The research questions themselves are not precise enough in their formulation. The results are insufficiently discussed. Please note, that the text references are based on the page and line number given in the manuscript. Abstract The abstract is well structured and easy to follow. To better understand your choice of research questions, a brief background section would be a helpful supplement. Background The introduction is well structured and nicely to read. Especially I enjoyed your summary on why medicine challenges student's wellbeing. Overall, I would encourage you to look more into literature that led you to the different parts of your research question (when, how, why, by whom …). Furthermore, it would be more coherent to integrate the pandemic-part of your introduction (page 5, line 7) into the remaining text. I think you wanted to demonstrate that exploring wellbeing in medical students became even more important during the pandemic. Maybe you could

	summarize this statement in the introduction and integrate it in your discussion. I would recommend you to be more consistent with your research questions. On page 3 (line 5), you do not name the 'by-whom-question' whereas on page 5 (line 37) you are adding the question 'what' without introducing it. To make the paper as clear as possible, I would formulate the hypotheses at the end of the introduction with the terms you were using on page 3, line 5. The sub-question "when wellbeing should be measured" is misleading, because the students do not answer when wellbeing sould be measured but when they can take 5 minutes in a certain period of time to measure their wellbeing. This is a big difference. Method The methods appear to be well founded for your exploratory approach. To better demonstrate the procedure of the study, a list of the questions from the survey would be recommended. Especially on page 12, line 51-54 just the top 4 determinants of wellbeing are listed - between which items the students could choose? Furthermore, I am wondering what definition of wellbeing the students received as part of the survey? Did their definition focus on explicitly on mental wellbeing? To make your sample description clearer for the reader, a reference to table 1 is recommended (page 4, line 54). Results The comparison with the National Comparator comes as a bit of a surprise. Perhaps you could introduce it in one sentence. It is not described with which intention the National Comparator is used. Also, only medical students were interviewed in their sample, while the National Comparator consists of different samples. No comma was used in the number n of ethnicity (p. 10, l. 21) and religion, which is inconsistent with the other values in the table. In your results section, 49.1% of 116 does not add up to an even number (page 11, line 37) – how is this possible? Discussion In order to keep the thread of the text, a short summary of the results is recommended at the beginning of the discussion. On page 14, line 19, you mention the number 17, which in itself is not very meaningful. To understand the significance of this number, it would be advisable to put it in relation to the total number of subjects giving answer to that question. Moreover, the number 17 is not to be found in your results section. Your ideas are nicely elaborated, but a discussion of quantitative findings is lacking. References The conclusion is well written and summarizes the most important points of the paper.

REVIEWER	Lo Moro, Giuseppina Università degli Studi di Torino
REVIEW RETURNED	12-Oct-2021

GENERAL COMMENTS	The paper is interesting and I think that the mixed-method approach is useful to explore how, why and when medical students think their wellbeing should be measured. I have made several

	suggestions that the Authors may want to consider in revising their paper. 1.Introduction 1.1 Line 38-39: 'Imposter syndrome' is not a full sentence, please rephrase 1.2 Second paragraph of introduction: I suggest some other papers to enrich challenges that medical student wellbeing: -Dyrbye LN, et al. Personal Life Events and Medical Student Burnout: A Multicenter Study. Acad Med 2006;81(4):374–84. pmid:16565189 -Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. Med Educ 2016;50(1):132–49. pmid:26695473 -Bert F, et al. (2020) Prevalence of depressive symptoms among Italian medical students: The multicentre cross-sectional "PRIMES" study. PLOS ONE 15(4): e0231845. https://doi.org/10.1371/journal.pone.0231845 -Fares J, et al. Extracurricular activities associated with stress and burnout in preclinical medical students. J Epidemiol Glob Health 2016;6(3):177–85. pmid:26644345 1.3 Pastoral support should be explained a little more in depth for an international readership. 2.Methods 2.1 The number of eligible students could be written in the Methods section. 2.2 Outcomes: Are also the demographics outcomes of the study? 2.3 Measurement: Please describe in more detail the measurement paragraph. 2.4 Statistical methods: You should describe the analyses that you've done, even if they're only descriptive analyses. 2.5 Semi-structured interviews: Please write the SRQR abbreviation meaning. Overall, all the abbrevations in the paper should be explained the first time they appear. 3.Results 3.1 Please explain what BMA is 3.2 Interview: it could be interesting to know how many students have discussed about the thematic you choose. For instance, how many students thought that exercise positively impact on wellbeing? 3.3 A total of 55 gave consent to do the qualitative part. Did the students who gave no consent have different sociodemographics? What were the sociodemographic characteristics of the 16 students who participated in the qualitative part? Were genders and minorities represented? 4.Discussion 4.1 Exercise is widely reported to condition medical student's wellbeing and mental health, here some references to enrich this part of the discussion: -Fares J, et al. Extracurricular activities associated with stress and burnout in preclinical medical students. J Epidemiol Glob Health 2016;6(3):177–85. pmid:26644345 -Bert F, et al. (2020) Prevalence of depressive symptoms among Italian medical students: The multicentre cross-sectional "PRIMES" study. PLOS ONE 15(4): e0231845. https://doi.org/10.1371/journal.pone.0231845 -Blake H, Stanulewicz N, Mcgill F. Predictors of physical activity and barriers to exercise in nursing and medical students. J Adv Nurs 2017;73(4):917–29. pmid:27731886

	4.2 Some references on the effectiveness of the self-care assessment could be helpful. 4.3 Limitations: a test could be used to test the statistical difference between your sample characteristics and the national sample. Moreover, the generalizability of the results should be discussed, as well as the limitations of the methods of sampling.

REVIEWER	Merlo, Lisa University of Florida, Psychiatry
REVIEW RETURNED	14-Oct-2021

GENERAL COMMENTS	Thank you for the opportunity to review “What medical students think about measurement of their wellbeing and implications for pastoral support: cross sectional survey and qualitative interviews.” This is an interesting topic, using a somewhat novel approach to obtain information to aid in efforts to better support medical students during their training. There are some issues that detract from the potential impact of the paper as currently written, which are outlined below. Title: It may be preferable to reword the title to be more descriptive of the specific study aims and findings. The results focus primarily on student preferences regarding timing and method of assessing their personal wellbeing, but that is not clear from the title. Note-- the term “pastoral support” may be confusing to American readers (and perhaps other international readers as well), as in the USA this term typically pertains to counseling provided by clergy. I was personally expecting to read a paper regarding the impact of involving religious leaders in medical student wellbeing programming. Please consider operationally defining this term. Introduction: There are a number of statements throughout the introduction that are confusing and/or merit further explanation. For example: • p. 4 lines 31-32: what does it mean to say that “Recruitment policy inclusivity has not yet caught up with content and support”? • line 37-38: incomplete sentence • lines 43-44: what is “the context of widening participation and graduate entry” and how is this related to financial stress? • Lines 53-54: What are “enabling services” and “occupational health” for medical students? • Line 56: what does is mean to be “medicine assessment literate”? • P. 5 line 7: what is an “open” learning environment? • The paragraph on p. 5 lines 7-21 is confusing as written Sample: The overall study response rate (9.5%) was extremely low, introducing a significant risk of self-selection bias within the sample. In addition, several questions had even lower response rates, further calling into question the generalizability of results. It would be useful to compare the demographics of the study sample to the population of students at the medical school as a whole. Similarly, the N of 118 did not meet the minimum sample size of 245-381 students that the authors indicated would be necessary for making assumptions generalizable to the national population of medical students. In addition, the participants are all drawn from a single medical school, which may significantly impact student views regarding the best method/timing of assessments, the appropriate confidantes (i.e., not all schools have PATs), and the utility of existing resources. This should be discussed as a

	limitation. There was very low participation by male students (even when compared to the national comparison data), which may result in biased results overall. Were any gender differences observed in the findings? Data collection: given that the study took place during the COVID-19 pandemic, more information is needed about how students' medical training was impacted. Were they attending in-person classes and participating in clinical rotations with real patients at the time of the study? Or were they attending virtual lessons from home and/or excluded from clinical settings? How might this have impacted student exposure to study invitations? How might it have impacted student interest in responding? How might student preferences regarding assessment of wellbeing change when not in the midst of the pandemic? Please review in Limitations and Discussion. Measurement: Much more information is needed about the study measure. What questions were asked? What options were offered for frequency of assessment, format of measurement, purpose of measurement, determinants of wellbeing, etc.? Which items were Likert response and which were free text? What were the response options for the 9-point scale (i.e., what did 1 and 9 correspond to?). Statistical methods: how were descriptive statistics computed? Were any inferential statistics considered? Gender differences? Differences by class year or pre-clinical vs. clinical students? Ethics: Are the numbers listed in parentheses telephone numbers for counseling services? If so, it may be best to exclude these from the manuscript so they do not receive inappropriate calls and/or in case the phone numbers change over time. Results: Information regarding cut points for "limited importance" etc. should be included in the Data Analysis section rather than the results. The result that half of students would spend at least 5 minutes per day completing a wellbeing assessment was very surprising, as this would not be the case among most medical students in the USA who complain of "survey fatigue." This finding should be discussed further—I would be interested in the authors' perspective regarding high rates of student interest (perhaps more information about this was gathered during the interviews?). Considering that a different n was used for most results, it would be helpful to include percentages for the respondents when reporting results. How were students meant to interpret "research," "national governance," "local governance," etc.? How was this question asked? Were these terms operationally defined for them? What other determinants of wellbeing were included in the list? Were there determinants that students felt were unimportant to assess? Discussion: The most important conclusion may be the statement that "students need to be taught that they will have deadlines and pressures throughout their careers and how to prioritise self-care" The section on "Suggested Self Care Assessments" was confusing. Who is suggesting this? On what basis are these

	suggestions being made? i.e., are these recommendations based on study findings or the authors' personal opinions? How would students provide quality assurance of pastoral support, and more importantly, of peer support? Given that students already report being stretched for time, when would they be expected to attending trainings to provide the support? Who would be tasked with serving as a peer supporter? How many students would they be expected to support? How/when/by whom would they be supervised/evaluated?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Miss Rebecca Erschens, University Medical Hospital Tuebingen, Internal Medicine

"Thank you for the opportunity to review the manuscript 'What medical students think about measurement of their wellbeing and implications for pastoral support: cross sectional survey and qualitative interviews'. This manuscript investigates medical student`s opinion on the measurement of wellbeing regarding the frequency of availability for measurement, the format, type and purpose of measurement and with whom wellbeing should be discussed. Medical students from different years of Southampton University took part in an online survey and semi-structured interviews."
Overall your manuscript is well structured and easy to follow. Your ideas regarding improving student well-being are well elaborated.

Thank you for your overall comments. We picked out the suggested areas for improvement below:

"The introduction does not adequately introduce the rationale for your choice of research questions"
We have re-written the penultimate sentence for the introduction to make this clearer.

"The research questions themselves are not precise enough in their formulation."
We have replaced the sentence summarising in a qualitative style the exploratory research questions with a list of more precise research questions.

"The results are insufficiently discussed"
This was difficult to address given the constraints of the author guideline, particularly as we have gone beyond description to suggested solutions, but an introductory sentence was added and more content under the subheading "Limitations".

Abstract
"The abstract is well structured and easy to follow. To better understand your choice of research questions, a brief background section would be a helpful supplement."
We agree that to have no brief background in the abstract feels odd, but the BMJ Open guidance is that no Background should appear in the abstract.

Background
"The introduction is well structured and nicely to read. Especially I enjoyed your summary on why medicine challenges student's wellbeing. Overall, I would encourage you to look more into literature that led you to the different parts of your research question (when, how, why, by whom …). "

Unfortunately, the author guidance on word count and accessibility to a general audience does not allow me to elaborate on this, but we added some text in brackets to the Outcomes section of the Methods to address your point.

Furthermore, it would be more coherent to integrate the pandemic-part of your introduction (page 5, line 7) into the remaining text. I think you wanted to demonstrate that exploring wellbeing in medical students became even more important during the pandemic. Maybe you could summarize this statement in the introduction and integrate it in your discussion.
To allow for further discussion of the results, it was not possible to do this within the constraints of the suggested number of Discussion paragraphs.

I would recommend you to be more consistent with your research questions. On page 3 (line 5), you do not name the 'by-whom-question' whereas on page 5 (line 37) you are adding the question 'what' without introducing it. To make the paper as clear as possible, I would formulate the hypotheses at the end of the introduction with the terms you were using on page 3, line 5.
The sub-question "when wellbeing should be measured" is misleading, because the students do not answer when wellbeing sould be measured but when they can take 5 minutes in a certain period of time to measure their wellbeing. This is a big difference.
We replaced the sentence summarising in a qualitative style the exploratory research questions with a list of the precise research questions.

1) How often medical students could give time for wellbeing measurement?
2) What format medical students want their wellbeing measured in?
3) For what purposes should medical student wellbeing be measured?
4) What type of wellbeing data should be measured for medical students?
5) Who do medical students feel comfortable speaking to about their wellbeing?
6) What determinants of wellbeing should be measured for medical students?

The list of Outcomes in the methods section has more detail added:

- Frequency of availability to spend 5 minutes measuring wellbeing
- Format of wellbeing measurement (e.g. survey, conversation)
- Purpose of wellbeing measurement (e.g. research, governance)
- Type of wellbeing measurement (e.g. evaluative, experienced, subjective, objective, quantitative or qualitative)
- Who to talk with about wellbeing at a 30 minute conversation
- Determinants of wellbeing that should be measured

Method
The methods appear to be well founded for your exploratory approach. To better demonstrate the procedure of the study, a list of the questions from the survey would be recommended. Especially on page 12, line 51-54 just the top 4 determinants of wellbeing are listed - between which items the students could choose?
We have provided BMJ Open with a supplementary file containing the survey and the DOI for the data for the study, which include all the questions and the answer options. I have also added further to the Measurement subsection of the Methods:
"Office for National Statistics 2011 census questions were utilised for demographics (age, sex, ethnicity, religion) and educational demographics about Year and Programme were used. Increasing increments of time were offered and one answer allowed for frequency of measurement. To assess agreement about the format, purpose and type of wellbeing measurement, 9-point Likert scales were used. Different roles and services were offered and students were able to select as many as they liked in answer to who they would feel comfortable talking to. A list of 47 determinants of wellbeing was offered, and multiple answers were allowed, students were asked to consider the survey burden

associated. No assumptions were made about student preferences, so free text answer options were available for all questions."

Furthermore, I am wondering what definition of wellbeing the students received as part of the survey? Did their definition focus on explicitly on mental wellbeing?
In the Bias subsection of the methods, we added: "A definition of wellbeing was not provided to allow individual interpretation and prevent bias."

To make your sample description clearer for the reader, a reference to table 1 is recommended (page 4, line 54).
We added "(n=1245)" this line to make the population clearer.

Results
The comparison with the National Comparator comes as a bit of a surprise. Perhaps you could introduce it in one sentence. It is not described with which intention the National Comparator is used.
This text has been added "The demographics of the 118 participants are shown in Table 1 alongside national UK Medical and Dental Student, or General Student demographics, to show the level of representation of the sample."

Also, only medical students were interviewed in their sample, while the National Comparator consists of different samples.
The national comparator data presented are the most relevant openly available.

No comma was used in the number n of ethnicity (p. 10, l. 21) and religion, which is inconsistent with the other values in the table.
We are not sure what comma was referred to here, as no commas were used in Table 1, but have given the n for religion a separate row to make it uniform with the other demographics as this is perhaps what was meant.
In your results section, 49.1% of 116 does not add up to an even number (page 11, line 37) – how is this possible?
57 participants chose increments of time that were at least once a day, or more frequent (57/116=0.49137), rounded to 49.1%

Discussion
In order to keep the thread of the text, a short summary of the results is recommended at the beginning of the discussion.

On page 14, line 19, you mention the number 17, which in itself is not very meaningful. To understand the significance of this number, it would be advisable to put it in relation to the total number of subjects giving answer to that question. Moreover, the number 17 is not to be found in your results section.
Thank you for spotting this typo. It now reads "This left 14.7% participants who could not give 5 minutes more often than once a month" In the results section and 14.7% in the discussion.

Your ideas are nicely elaborated, but a discussion of quantitative findings is lacking.
Thank you, further content has been added to the Limitations subsection: "The response rate for the survey was low (9.5%) but reflected the level of engagement seen with a Wellbeing Workshop (4%), offered during the UKs second lockdown, to the same population and the response rates of <10% commonly seen for online surveys (65). The study was undertaken at one Medical School and lacked the power to enable national inferences, although the demographic characteristics of this sample are broadly similar to those of medical students nationally. As in most survey research fewer participants

identifying as male responded (66), again highlighting the need for assessment to engage male medical students in wellbeing measurement."

References
The conclusion is well written and summarizes the most important points of the paper.
Thank you

Reviewer: 2
Dr. Giuseppina Lo Moro, Università degli Studi di Torino

Comments to the Author:
The paper is interesting and I think that the mixed-method approach is useful to explore how, why and when medical students think their wellbeing should be measured. I have made several suggestions that the Authors may want to consider in revising their paper.
Thank you for your summary comments.
1.Introduction
1.1 Line 38-39: 'Imposter syndrome' is not a full sentence, please rephrase
Thank you for spotting this it now reads "'Imposter syndrome' (14) and the culture of competition,"

1.2 Second paragraph of introduction: I suggest some other papers to enrich challenges that medical student wellbeing:

-Dyrbye LN, et al. Personal Life Events and Medical Student Burnout: A Multicenter Study. Acad Med 2006;81(4):374–84. pmid:16565189 -Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. Med Educ 2016;50(1):132–49. pmid:26695473

Now referenced in the subsection "Demographics beyond gender…" in the discussion "This might help mitigate the increased risk of burnout seen with an increasing number of negative life events in medical students (65)."

Bert F, et al. (2020) Prevalence of depressive symptoms among Italian medical students: The multicentre cross-sectional "PRIMES" study. PLOS ONE 15(4): e0231845. https://doi.org/10.1371/journal.pone.0231845
We have added both these references into the Introduction: "The focus tends to be on pathology, when problems have already occurred, such as failed assessments, or mental health diagnoses (29, 30)"

-Fares J, et al. Extracurricular activities associated with stress and burnout in preclinical medical students. J Epidemiol Glob Health 2016;6(3):177–85. pmid:26644345
Included in the discussion now in the Subsection on "Self care…": Exercise …has been shown to reduce stress, anxiety and depression in medical students (53)."

1.3 Pastoral support should be explained a little more in depth for an international readership.
We have removed the word 'pastoral', as this is a specific term, often with religious connotations.

2.Methods
2.1 The number of eligible students could be written in the Methods section.
We have added (n=1245) to the sample subsection.

2.2 Outcomes: Are also the demographics outcomes of the study?
Good point. We have moved the Demographics into the Data Collection section.

2.3 Measurement: Please describe in more detail the measurement paragraph.
We address each question now in the measurement paragraph: “Office for National Statistics 2011 census questions were utilised for demographics (age, sex, ethnicity, religion) and educational demographics about Year and Programme were used. Increasing increments of time were offered and one answer allowed for frequency of measurement. To assess agreement about the format, purpose and type of wellbeing measurement, 9-point Likert scales were used. Different roles and services were offered and students were able to select as many as they liked in answer to who they would feel comfortable talking to. A list of 47 determinants of wellbeing was offered, and multiple answers were allowed, students were asked to consider the survey burden associated. No assumptions were made about student preferences, so free text answer options were available for all questions.”

2.4 Statistical methods: You should describe the analyses that you’ve done, even if they’re only descriptive analyses.
This sentence has been added “The numbers of participants that selected an option are reported as percentages of those that responded to the question, where a single answer was permitted.”

2.5 Semi-structured interviews: Please write the SRQR abbreviation meaning. Overall, all the abbrevations in the paper should be explained the first time they appear.
We corrected this omission “Reported using Standards for Reporting Qualitative Research (SRQR) recommendations”

3.Results
3.1 Please explain what BMA is
Thank you for highlighting this it now reads “the British Medical Association (BMA, the UK Doctor Trade Union)”
3.2 Interview: it could be interesting to know how many students have discussed about the thematic you choose. For instance, how many students thought that exercise positively impact on wellbeing?
It was decided a priori that quantitative analysis would not be undertaken, that inductive analysis would be used.

3.3 A total of 55 gave consent to do the qualitative part. Did the students who gave no consent have different sociodemographics? What were the sociodemographic characteristics of the 16 students who participated in the qualitative part? Were genders and minorities represented?
Sadly,we are not ‘allowed’ any more tables or figures to present the demographics of this sub-group.

4.Discussion
4.1 Exercise is widely reported to condition medical student’s wellbeing and mental health, here some references to enrich this part of the discussion:
-Fares J, et al. Extracurricular activities associated with stress and burnout in preclinical medical students. J Epidemiol Glob Health 2016;6(3):177–85. pmid:26644345 –
Included in the Discussion, in the sub-section on “Self care…”: Exercise …has been shown to reduce stress, anxiety and depression in medical students (53).”

Bert F, et al. (2020) Prevalence of depressive symptoms among Italian medical students: The multicentre cross-sectional “PRIMES” study. PLOS ONE 15(4): e0231845. https://doi.org/10.1371/journal.pone.0231845
This reference has been added to the Introduction, there was not an appropriate place to also include it in the Discussion.

-Blake H, Stanulewicz N, Mcgill F. Predictors of physical activity and barriers to exercise in nursing and medical students. J Adv Nurs 2017;73(4):917–29. pmid:27731886

This reference has been added to the “Self care needs ..” subsection of the discussion.

4.2 Some references on the effectiveness of the self-care assessment could be helpful.
We believe this is novel, and no work has been published on this before.
4.3 Limitations: a test could be used to test the statistical difference between your sample characteristics and the national sample. Moreover, the generalizability of the results should be discussed, as well as the limitations of the methods of sampling.
We expanded on the limitations: The response rate for the survey was low (9.5%), but reflected the level of engagement seen with a Wellbeing Workshop (4%), offered during the UKs second lockdown, to the same population and the response rates of <10% commonly seen for online surveys (68). The study was undertaken at one Medical School and lacked the power to enable national inferences, although the demographic characteristics of this sample are broadly similar to those of medical students nationally. As in most survey research fewer participants identifying as male responded (69), again highlighting the need for assessment to engage male medical students in wellbeing measurement.
In the statistical methods subsection the following has been added in addition to an extra column in the Demographics table “The difference between the demographics of this sample and the closest available national comparable population was assessed for significance through calculating the standard error and confidence intervals for the difference.”

Reviewer: 3
Dr. Lisa Merlo, University of Florida

Thank you for the opportunity to review “What medical students think about measurement of their wellbeing and implications for pastoral support: cross sectional survey and qualitative interviews.” This is an interesting topic, using a somewhat novel approach to obtain information to aid in efforts to better support medical students during their training. There are some issues that detract from the potential impact of the paper as currently written, which are outlined below.
Thank you for your opening comments.

Title: It may be preferable to reword the title to be more descriptive of the specific study aims and findings. The results focus primarily on student preferences regarding timing and method of assessing their personal wellbeing, but that is not clear from the title.
Note-- the term “pastoral support” may be confusing to American readers (and perhaps other international readers as well), as in the USA this term typically pertains to counseling provided by clergy. I was personally expecting to read a paper regarding the impact of involving religious leaders in medical student wellbeing programming. Please consider operationally defining this term.

We have changed the title to make the study more internationally accessible as you suggest “What medical students think about measurement of their wellbeing and the implications for Universities: cross sectional survey and qualitative interviews” I have not used the word preferences as this suggests a solely subjective opinion, when students interviewed had thought processes that went beyond this to encompass policy, system constraints and other students. Since more aspects of wellbeing measurement than timing and method are captured, we have not added these terms to the title.

Introduction: There are a number of statements throughout the introduction that are confusing and/or merit further explanation. For example:
- p. 4 lines 31-32: what does it mean to say that “Recruitment policy inclusivity has not yet caught up with content and support”?

You are absolutely correct, this sentence was confusing and has been re-written “Recruitment policy inclusivity, in the form of Graduate entry and Widening participation programmes, has not yet been matched by equally accessible, diverse and inclusive content and support.”

- line 37-38: incomplete sentence

Thank you for spotting this it now reads “‘Imposter syndrome’ (14) and the culture of competition,”

- lines 43-44: what is “the context of widening participation and graduate entry” and how is this related to financial stress?

I have clarified this further “The length of the course leads to more students facing financial stressors, especially in the context of widening participation, where students may come from lower income families and graduate entry where students already have significant debt (16).”

- Lines 53-54: What are “enabling services” and “occupational health” for medical students?

I have explained these terms further “The need to provide not only enabling services (20) that support students with learning difficulties and disabilities, but also occupational health (21), which promotes and ensures fitness to work, for example, is unique to healthcare students.”

- Line 56: what does is mean to be “medicine assessment literate”?

A reference is provided that explains this term, which means to know about the assessments medical students have as part of their degree programme “Generic university wellbeing support provisions are often not medicine assessment literate (22),”

- P. 5 line 7: what is an “open” learning environment?

A reference is provided that explains this term, which means an environment in which students are encouraged to share experiences and have freedom of speech, and should expect to be kept informed and to receive only constructive feedback: “All students require a safe learning environment (19) that is open (24)”. Unfortunately, the word count does not allow elaboration in the text.

- The paragraph on p. 5 lines 7-21 is confusing as written

This paragraph has been re-written: “All students require a safe learning environment (19) that is open (24), but these most basic needs have been hard to meet at Medical School in the context of a pandemic. Necessary, abrupt, changes such as the move to online learning were shown to be anxiety provoking (25) for practical content such as anatomy. There was uncertainty about the loading of practical skills into periods when lockdowns were lifted, and concerns about whether online assessment would allow progression and graduation (26). The loss of face to face interaction lead to a loss of peer support on how to deal with uncertainty (27). Final year medical students still had to attend placements that brought them into direct contact with the Covid 19 virus (28). In these ways the wellbeing of medical students was particularly impacted by the pandemic.”

Sample: The overall study response rate (9.5%) was extremely low, introducing a significant risk of self-selection bias within the sample. In addition, several questions had even lower response rates, further calling into question the generalizability of results. It would be useful to compare the demographics of the study sample to the population of students at the medical school as a whole.

We agree the response rate was low and have elaborated on this further in the Limitations section of the discussion “The response rate for the survey was low (9.5%), but reflected the level of engagement seen with a Wellbeing Workshop (4%), offered during the UKs second lockdown, to the same population and the response rates of <10% commonly seen for online surveys”

Similarly, the N of 118 did not meet the sample size of 245-381 students that the authors indicated would be necessary for making assumptions generalizable to the national population of medical students. In addition, the participants are all drawn from a single medical school, which may significantly impact student views regarding the best method/timing of assessments, the appropriate confidantes (i.e., not all schools have PATs), and the utility of existing resources.

Indeed, different systems and therefore findings might arise from studies undertaken at other medical schools, but the broad concepts that students should be asked about measurement and that choice

and quality assurance are important to students are I think likely to be universal given the high percentages that thought this was critically important.

This should be discussed as a limitation. There was very low participation by male students (even when compared to the national comparison data), which may result in biased results overall. Were any gender differences observed in the findings?

We added to the Limitations subsection to address this point “As in most survey research fewer participants identifying as male responded, again highlighting the need for assessment to engage male medical students in wellbeing measurement.” We have not undertaken subgroup analysis due to the lack of power.

Data collection: given that the study took place during the COVID-19 pandemic, more information is needed about how students’ medical training was impacted. Were they attending in-person classes and participating in clinical rotations with real patients at the time of the study? Or were they attending virtual lessons from home and/or excluded from clinical settings? How might this have impacted student exposure to study invitations? How might it have impacted student interest in responding? How might student preferences regarding assessment of wellbeing change when not in the midst of the pandemic? Please review in Limitations and Discussion.

We added to the Context subsection of the methods “in the Covid 19 pandemic. During this period students in the final year attended clinical placements, while all other years had a period in which all their training was delivered online only.”
We added to the Discussion section as you suggest “The response rate for the survey was low (9.5%) but reflected the level of engagement seen with a Wellbeing Workshop (4%), offered during the UKs second lockdown, to the same population and the response rates of <10% commonly seen for online surveys (69).”

Measurement: Much more information is needed about the study measure. What questions were asked? What options were offered for frequency of assessment, format of measurement, purpose of measurement, determinants of wellbeing, etc.? Which items were Likert response and which were free text? What were the response options for the 9-point scale (i.e., what did 1 and 9 correspond to?).
The measurement subsection of the methods has been added to, to cover your comments:
“Office for National Statistics 2011 census questions were utilised for demographics (age, sex, ethnicity, religion) and educational demographics about Year and Programme were used. Increasing increments of time were offered and one answer allowed for frequency of measurement. To assess agreement about the format, purpose and type of wellbeing measurement, 9-point Likert scales were used. On the 9-point Likert scale the boundaries were categorised as follows: Limited Importance = 1-3, Important = 4-6, Critically important = 7-9. Different roles and services were offered and students were able to select as many as they liked in answer to who they would feel comfortable talking to. A list of 47 determinants of wellbeing was offered, and multiple answers were allowed, students were asked to consider the survey burden associated. No assumptions were made about student preferences, so free text answer options were available for all questions.”

Statistical methods: how were descriptive statistics computed? Were any inferential statistics considered? Gender differences? Differences by class year or pre-clinical vs. clinical students? Unfortunately, there was insufficient space to include the programme and year demographics.

Ethics: Are the numbers listed in parentheses telephone numbers for counseling services? If so, it may be best to exclude these from the manuscript so they do not receive inappropriate calls and/or in case the phone numbers change over time.

Good point, we removed these.

Results: Information regarding cut points for “limited importance” etc. should be included in the Data Analysis section rather than the results.
The cut off points are mentioned again in the table legends, to assist the reader.

The result that half of students would spend at least 5 minutes per day completing a wellbeing assessment was very surprising, as this would not be the case among most medical students in the USA who complain of “survey fatigue.” This finding should be discussed further—I would be interested in the authors’ perspective regarding high rates of student interest (perhaps more information about this was gathered during the interviews?).
As wellbeing measurement frequency is a novel area of research in medical students I cannot compare what we found in this sample to others to explain the findings. Students did mention feedback fatigue in the interviews.

Considering that a different n was used for most results, it would be helpful to include percentages for the respondents when reporting results.
We reported percentages for all the questions where one answer was allowed.

How were students meant to interpret “research,” “national governance,” “local governance,” etc.? How was this question asked? Were these terms operationally defined for them?
What other determinants of wellbeing were included in the list? Were there determinants that students felt were unimportant to assess?
A supplementary pdf of the survey, which contains the answers to these questions has been uploaded.

Discussion: The most important conclusion may be the statement that “students need to be taught that they will have deadlines and pressures throughout their careers and how to prioritise self-care”
The section on “Suggested Self Care Assessments” was confusing. Who is suggesting this? On what basis are these suggestions being made? i.e., are these recommendations based on study findings or the authors’ personal opinions?
I have cut and updated this section for clarity: “Suggested Self Care Assessment:
Objective Structured Clinical Examinations (OSCEs) that allow medical students to be the person running a Health Education England recommended wellbeing check-in (64), or attending one, with learning outcomes around communication skills, signposting and boundary setting, or self-monitoring and self-care respectively. These would facilitate the culture of peer support students wanted in this study, as well as meet the UK recommendation for wellbeing check-ins 2 weeks into every placement (64). These could be undertaken with other health and social care students.”

How would students provide quality assurance of pastoral support, and more importantly, of peer support? Given that students already report being stretched for time, when would they be expected to attending trainings to provide the support? Who would be tasked with serving as a peer supporter? How many students would they be expected to support? How/when/by whom would they be supervised/evaluated?
I have added this sentence to address some of your questions “There is an appetite among a group of the students at the University of Southampton to deliver a Peer Support programme, with recruitment and training underway.”

VERSION 2 – REVIEW

REVIEWER	Erschens, Rebecca University Medical Hospital Tuebingen, Internal Medicine, Department of Psychosomatic Medicine and Psychotherapy
REVIEW RETURNED	31-Jan-2022

GENERAL COMMENTS	Dear Authors, Thank you for the revision of the points raised by the reviewers, which I consider to have been comprehensively addressed.

REVIEWER	Lo Moro, Giuseppina Università degli Studi di Torino
REVIEW RETURNED	25-Nov-2021

GENERAL COMMENTS	The Authors successfully addressed all my previous comments.

REVIEWER	Merlo, Lisa University of Florida, Psychiatry
REVIEW RETURNED	05-Dec-2021

GENERAL COMMENTS	Thank you for the opportunity to review the revised version of this paper. The authors have made some significant improvements; however, there are remaining concerns that detract from its potential to contribute to the extant literature. The primary concern is the very small and likely non-representative nature of the study sample. Although the pool of eligible participants was 1245, the highest number of responses obtained on any item was n = 118 (i.e., 9.5% response rate), with several results reflecting even lower rates (i.e., n = 95 or 7.6% response rate). This number did not come close to achieving the authors' pre-determined requirement for a representative sample (i.e., N = 245-381). In addition, comparison of the sample to national norms indicated significant differences in the gender, age, and racial/cultural background of the two groups. This is particularly concerning because the study purports to assess how/when/why/by whom students should provide feedback on their wellbeing. Not surprisingly, the minority of students who volunteered to complete a self-report survey regarding assessment of wellbeing on their own time reported being interested in and willing to voluntarily complete a self-report survey assessing their wellbeing on their own time. It is unlikely that their responses could be generalized to the >90% of students who chose not to complete this survey. This may be a fatal flaw, and requires that the authors address it fully in the Limitations section, along with calls for future research that includes larger, more representative samples. On a related note, the authors attempt to make a case that the low rate of response is similar to an even lower rate of participation in a voluntary wellness program, but this further emphasizes the problem—namely, current efforts to target medical student wellbeing are not reaching the majority of students. It is imperative to find a way to reach those who do not respond to optional wellness-related opportunities (particularly as there is some

	evidence that those who do respond tend to be more invested in their wellbeing, and thereby also display healthier habits). For the qualitative analysis, only 16 participants were included (16/1245 = 1.3%) and there is no indication of true saturation being achieved in the results. This corresponds to the second major concern-- namely, the relative lack of information regarding the qualitative analyses performed in this study. It would be helpful to explain how the themes were identified and by whom—was any established methodology used? Was there any measure of interrater reliability? How were the representative quotes chosen? Finally, the authors declined to address multiple reviewer comments on the basis of lacking adequate space in the manuscript to do so. However, they could likely save a number of words by refining the existing methods and discussion sections, as well as by consolidating the presentation of the qualitative results. This would allow them to more clearly delineate their analyses and results, which would likely be more valuable to the reader.