



SYSTEMATIC REVIEW

A systematic review of effective quality feedback measurement tools used in clinical skills assessment [version 1; peer review: 1 approved, 1 approved with reservations]

Akram Alsaifi ^{1,2}, Davina Li Xin Ling¹, Micheál Newell¹, Thomas Kropmans¹

¹College of Medicine, Nursing and Health Sciences – School of Medicine, National University of Ireland, Galway, Galway, Galway. Co, H91 V4AY, Ireland

²Department of Medical Education, College of Medicine, Taif University, Saudi Arabia, P.O Box 11099, Taif 21944, Saudi Arabia

V1 First published: 25 Feb 2022, 12:11
<https://doi.org/10.12688/mep.18940.1>

Latest published: 19 Jun 2023, 12:11
<https://doi.org/10.12688/mep.18940.2>

Abstract

Background:

Objective Structured Clinical Examination (OSCE) is a valid tool to assess the clinical skills of medical students. Feedback after OSCE is essential for student improvement and safe clinical practice. Many examiners do not provide helpful or insightful feedback in the text space provided after OSCE stations, which may adversely affect learning outcomes. The aim of this systematic review was to identify the best determinants for quality written feedback in the field of medicine.

Methods:

PubMed, Medline, Embase, CINAHL, Scopus, and Web of Science were searched for relevant literature up to February 2021. We included studies that described the quality of good/effective feedback in clinical skills assessment in the field of medicine. Four independent reviewers extracted determinants used to assess the quality of written feedback. The percentage agreement and kappa coefficients were calculated for each determinant. The ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool was used to assess the risk of bias.

Results:

14 studies were included in this systematic review. 10 determinants were identified for assessing feedback. The determinants with the highest agreement among reviewers were specific, described gap, balanced, constructive and behavioural; with kappa values of 0.79, 0.45, 0.33, 0.33 and 0.26 respectively. All other determinants had low agreement (kappa values below 0.22) indicating that even though they have been used in the literature, they might not be applicable for good quality feedback. The risk of bias was low or moderate overall.

Conclusions:

This work suggests that good quality written feedback should be specific, balanced, and constructive in nature, and should describe the

Open Peer Review

Approval Status

1

2

version 2

(revision)
19 Jun 2023

version 1

25 Feb 2022



[view](#)



[view](#)

1. **Reina Abraham**, University of KwaZulu-Natal, Durban, South Africa

2. **Subha Ramani** , Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Any reports and responses or comments on the article can be found at the end of the article.

gap in student learning as well as observed behavioural actions in the exams. Integrating these determinants in OSCE assessment will help guide and support educators for providing effective feedback for the learner.

Keywords

Objective Structured Clinical Examination (OSCE), effective feedback, measurement, quality, determinant, Kappa.

Corresponding author: Akram Alshafi (a.alsahafi1@nuigalway.ie)

Author roles: **Alshafi A:** Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Ling DLX:** Formal Analysis, Methodology; **Newell M:** Formal Analysis, Supervision, Writing – Review & Editing; **Kropmans T:** Conceptualization, Formal Analysis, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2022 Alshafi A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Alshafi A, Ling DLX, Newell M and Kropmans T. **A systematic review of effective quality feedback measurement tools used in clinical skills assessment [version 1; peer review: 1 approved, 1 approved with reservations]** MedEdPublish 2022, 12:11 <https://doi.org/10.12688/mep.18940.1>

First published: 25 Feb 2022, 12:11 <https://doi.org/10.12688/mep.18940.1>

Introduction

During their undergraduate education, medical and health sciences students are subjected to numerous clinical practical assessments in order to evaluate their performance¹. Feedback is a fundamental and important learning tool in medical education². Good and effective feedback assists students in accomplishing both learning and professional development, enhancing student motivation and satisfaction³⁻⁵. The Objective Structured Clinical Examination (OSCE) is a commonly utilized clinical skills assessment in medical and health sciences that has a positive impact on medical education⁶. OSCE is useful in the field of medicine for evaluating student performance for a variety of reasons; the OSCE will simulate the realities of clinical practice, enhancing students' confidence and ensuring safe clinical practice, with assessment based on objective determinants⁷⁻¹⁰. The OSCE is a valid and reliable assessment tool in a variety of fields, including medicine^{9,11-15}.

During the OSCE, examiners are requested to input the students' observed marks on score sheets (without knowing total marks) and can also provide their professional opinion on students' performance using the Global Rating Scale (GRS) (Fail, Borderline, Pass, Good, Excellent) based on experience^{16,17}. Previous research has shown a mismatch between observed Marks and GRS¹⁸. For example, the student may score a high result in the observation section, but receive a 'fail' for their Global Rating Score and potentially *vice versa*.

However, written feedback for OSCE is optional, despite previous research showing it to have a significant positive impact on student's learning outcomes^{19,20}. It is argued that many examiners may find it difficult to offer detailed or useful written feedback during OSCE evaluation^{21,22} due to time constraints as well as a 'judgement dilemma' of not knowing how much feedback or the type of feedback to give²³.

Even if written feedback is provided, to date there is no recognised objective measurement scoring tool that measures the quality of written feedback from the OSCE. Measuring the quality of written feedback will help examiners to improve their skills in feedback delivery, as well as encourage students to understand the OSCE marks they received and where they can improve in the future. In other fields of education, feedback quality measurement tools are used effectively to improve the quality of written feedback for students²⁴⁻²⁶. In order to develop such a tool, it is necessary to identify the determinants that result in effective written feedback. The main objectives of this systematic review are to identify and evaluate studies that have measured written feedback quality.

Methods

Study design

This a comprehensive systematic review to identify the most relevant determinants that describe good and effective written feedback in the field of medicine. Measurement tools that measure feedback quality both quantitatively and qualitatively will be included.

Search strategy and publication sources

CINHAL, PubMed, Medline, Embase, Scopus and Web of Science were searched for relevant studies published from January

2010 until February 2021. We used the following keywords: "OSCE", "objective structured clinical examination*", "medical student*", "medical education", "clinical skill*", "clinical setting." AND "formative feedback", "constructive feedback", "effective feedback", "qualitative feedback", AND "quality", "scoring", "measurement*/measuring", "assessment*/assessing".

We sought assistance from a university librarian to enhance our search strategy. The reference section of initially selected studies was also searched thoroughly for any additional relevant publications. A bibliographical database was created to store and manage the references.

Selection of articles

Each author independently screened retrieved articles against inclusion and exclusion criteria, and as a team agreed on the included studies. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines²⁷, only studies written in the English language, published in the last 10 years in the field of medicine were included. Both quantitative and qualitative studies were included. Studies were included if they described the quality of good/effective feedback in clinical skills assessment, or attempted to evaluate the quality of feedback in clinical skills assessment (e.g., OSCE), or described the quality of written feedback by enumerating determinants of effective feedback involving undergraduate students and postgraduate trainees. Exclusion criteria included papers not written in English language, case reports, 'grey literature' (which includes conference proceeding studies) and commentaries. Due to different cognitive demands and scopes of practice, publications relating to nursing, paramedical disciplines, pharmacy, and veterinary education were excluded (Figure 1). Reference lists of included studies were also explored to identify any additional studies.

Data extraction

Four independent reviewers (A.A, D.L, M.N and T.K) identified and extracted the determinants used to evaluate written feedback from the included studies (Table 1). An accumulated list of identified determinants and their respective definitions was compiled. Each reviewer then scored each of the included studies against the accumulated list of determinants. If the determinant was not mentioned by name specifically, it may have been used by its definition. A positive (+) sign indicated 'matching with determinant definition' while a negative (-) sign indicated 'no match with the determinant definition'.

Statistical analysis

Level of agreement scores (%) and Kappa coefficients between the reviewers were calculated for each determinant included. Both level of agreement and kappa were measured using an online calculator (<http://justusrandolph.net/kappa/>)²⁸. Percentage agreement calculates agreement by chance which is corrected for by calculating kappa. The average Kappa coefficients were interpreted as follows: <0 indicates no agreement, 0.01–0.20 indicates slight or poor agreement, 0.21–0.40 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–1.00 indicates almost perfect agreement²⁹. Determinants with the

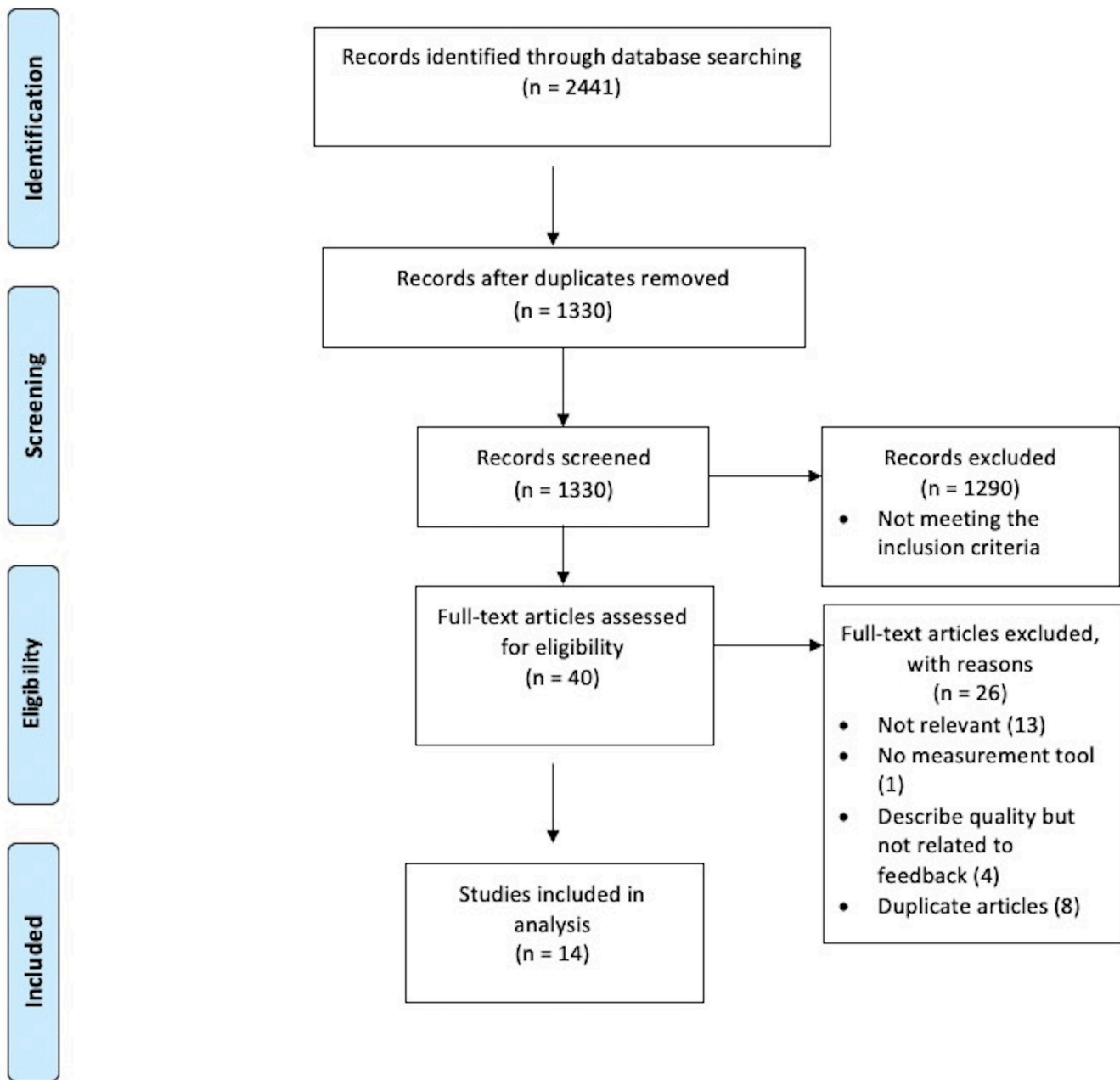


Figure 1. Flow diagram showing study selection process.

highest Kappa were identified as being most useful for providing written feedback for OSCE. In addition, included studies were assessed to identify which studies were best for measuring the quality of feedback.

Risk of bias and certainty assessment

Two independent reviewers used the ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool to assess the risk of bias of each included study (Table 3). Confounding,

selection, classification, intervention, missing data, measurement, and reporting were all checked for bias. The ROBINS-I was used to assess the certainty in the body of evidence in the context of GRADE's (Grading of Recommendations, Assessment, Development and Evaluations) approach^{30,31}.

When there were any conflicts, the entire review team was consulted, and the disagreements were then addressed by consensus.

Table 1. The 10 determinants of feedback quality measurement identified.

	Determinant of feedback measurement	General description
1	Specific	Detailed information of what was done well or poorly.
2	Balanced	Contains both positive and negative comments.
3	Behavioural	Observed action in exam (not personal).
4	Timely	Given immediately after assessment is completed.
5	Constructive	Supportive feedback identifying a solution to area of weakness they may have.
6	Quantifiable	Feedback that can be used to develop detailed statistical data.
7	Focused	Feedback that is given around key results.
8	Described task	Focuses the knowledge and skills associated with a task: sufficient or insufficient.
9	Described gap	Detailed about what is missing in the task.
10	Described action plan	Detailed plan of action needed to reach one or more goals.

Results

Search results

The initial search yielded 2441 studies (Figure 1). After the duplicates were removed, 1330 studies remained. 1290 studies were found to be irrelevant to the main topic after scanning the title and abstract. The 40 remaining articles were thoroughly evaluated by reading the full text. A further 26 articles^{2,7,21–23,32–44} were removed leaving 14 studies for inclusion in this systematic review (Figure 1).

Content analysis

Of the 14 studies included, 7 were conducted in the United States^{23,45–50}, 2 each in the United Kingdom^{51,52}, and Canada^{53,54}, and 1 each in the Netherlands⁵⁵, Switzerland⁵⁶, and South Africa¹⁹ respectively. Half of the studies involved medical students (undergraduate setting)^{46,48,49,51,53–55}, and the other half involved medical residents (postgraduate setting)^{19,45,47,50,52,56,57}. 10 of the 14 studies used a scoring system or systematic framework for evaluating the quality of written feedback^{19,45,48–54,57}, while the other 4 studies had no scoring system^{46,47,55,56}. Only 2 studies were conducted in multiple institution contexts^{48,50}, while 12 were conducted in a single institution^{19,45–47,49,51–57} setting.

Determinants identified

A total of 10 determinants to assess the quality of written feedback were identified from the combined 14 studies (Table 1).

Each reviewer then scored each of the 14 included studies against the accumulated list of determinants (Table 2).

The number of determinants identified in the individual studies ranged from 7 to 10 respectively. The determinants with the highest agreement (kappa values) among reviewers were Specific (0.79 - substantial agreement), Described gap (0.45 moderate

agreement), Balanced (0.33 - fair agreement), Constructive (0.33 - fair agreement), and Behavioural (0.26 - fair agreement) respectively. All other determinants had low agreement (kappa values below 0.21 - slight or poor agreement) indicating that even though they have been used in the literature, they might not be applicable for good quality feedback. The identified determinants with highest level of agreement among reviewers were included in seven of the ten studies of which the study by Abraham *et al.*¹⁹ had the highest level of agreement^{19,28,32,33,35,38,39}. (Table 2).

Risk of bias

We utilized the ROBINS-I score method to analyse bias across confounding bias, selection bias, classifications bias, intervention bias, bias due to missing data, measurement bias, and reporting bias to assess the possible risk of bias (Table 3). Almost all the studies included had low confounding, selection, and measurement biases. The overall risk of bias was low to moderate for the included studies, which is understandable considering the nonrandomized character of the research and dependence on self-reporting measures. The remaining post-intervention biases were variable, ranging from mild to moderate.

Certainty in body of evidence

In systematic reviews, the GRADE working group has created a widely-accepted approach to evaluate the certainty of a body of evidence-based on a four-level system: high, moderate, low and very low. The current GRADE strategy for a body of evidence linked to interventions starts by categorizing studies into one of two groups: randomized controlled trials (RCT) or observational studies (also non-randomized studies, or NRS). The body of evidence begins with high certainty if the relevant research is randomized trials. The body of evidence begins with a low level of certainty if the relevant study is observational^{31,58}.

Table 2. Scoring by reviewers for determinants of quality of written feedback.

	Year	Specific	Balanced	Behavioral	Timely	constructive	Quantifiable	Focused	Describe Task	Describe gap	Describe Action plan	Overall agreement % (Studies)	Kappa
Camarata, T. <i>et al.</i> ⁴⁵	2020	++++	++++	+++	++++	++++	++ ++	+++	+++	+++	+++	66.7	0.33
Tekian, A. <i>et al.</i> ⁴⁶	2019	++++	+++	++++	----	+++	+++	+++	+++	----	----	68.3	0.37
Tomiak, A. <i>et al.</i> ⁵³	2019	++++	----	++++	++++	++++	+++	++++	++++	+++	+++	78.3	0.57
Page, M. <i>et al.</i> ⁵¹	2019	++++	+++	++++	+++	++++	++ ++	+++	+++	----	+++	63.3	0.27
Abraham, R. <i>et al.</i> ¹⁹	2019	++++	+++	++++	+++	+++	++ ++	+++	++++	++++	++++	71.7	0.43
Dallaghan, G. <i>et al.</i> ⁴⁷	2018	++++	+++	++++	+++	+++	+++	+++	+++	----	+++	58.3	0.17
Karim, A. <i>et al.</i> ⁴⁸	2017	++++	+++	++++	+++	+++	+++	+++	+++	----	+++	53.3	0.07
Gullbas, L. <i>et al.</i> ⁵⁷	2016	+++	+++	++++	+++	+++	+++	+++	----	----	+++	65.0	0.30
Junod, N. <i>et al.</i> ⁵⁶	2016	+++	++++	++++	+++	+++	+++	+++	+++	+++	+++	53.3	0.07
Nesbitt, A. <i>et al.</i> ⁵²	2014	+++	+++	+++	+++	+++	----	+++	+++	++++	+++	51.7	0.03
Jackson, J. <i>et al.</i> ⁴⁹	2014	++++	++++	++++	+++	++++	+++	+++	+++	+++	+++	68.3	0.37
Gauthier, S. <i>et al.</i> ⁵⁴	2014	++++	----	+++	----	----	++ ++	+++	++++	++++	++++	88.3	0.77
Pelgrim, E. <i>et al.</i> ⁵⁵	2012	++++	+++	+++	+++	+++	+++	+++	+++	+++	++++	50.0	0.00
Canavan, C. <i>et al.</i> ⁵⁰	2010	++++	+++	++++	+++	+++	+++	+++	+++	+++	+++	56.7	0.13
Overall agreement % (determinants)		89.3	66.7	66.7	53.6	61.9	64.3	47.6	56	72.6	58.3		

Discussion

The aim of this systematic review was to evaluate studies that have measured the quality of written feedback for clinical exams and identify which determinants should be used to provide quality written feedback. Improving the quality of written feedback for students in the field of medicine will improve student performance.

Four independent researchers critically appraised 14 studies using 10 identified determinants. The five determinants with the highest Kappa values were: (1) Specific: tutors should include details in the comment section about what the student has done in the clinical exam (Kappa 0.79); (2) Described gap:

the comment should include points about what was missing in the task (Kappa 0.45);⁵⁹ Balanced: the comment should include both positive and negative statements about the student's performance (Kappa 0.33); (4) Constructive feedback: the comment should identify an area of improvement and give a solution to the student (Kappa 0.33); and (5) Behavioural: the comment should include observed (not personal) action in the exam (Kappa 0.26). The other five determinants were conversely deemed to be lacking in agreement, showing some form of confusion and complexity amongst the reviewers in its ascertainment. Hence, these determinants may not be considered to be a good qualitative measurement element of feedback quality.

Table 3. Results risk of bias assessment for individual studies using ROBINS-I methods.

Study		Pre-intervention		At-intervention	Post-intervention			Overall risk of bias	
First author	Year	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended Interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall risk of bias
Camarata ⁴⁵	2020	low	low	low	moderate	low	low	low	moderate
Tekian ⁴⁶	2019	low	low	low	low	low	low	low	low
Tomiak ³³	2019	low	low	low	moderate	low	moderate	moderate	moderate
Page ⁵¹	2019	low	low	low	low	low	low	low	low
Abraham ¹⁹	2019	low	low	low	low	low	low	low	low
Dallaghan ⁴⁷	2018	Low	moderate	Low	low	moderate	low	moderate	moderate
Karim ⁴⁸	2017	Low	low	Low	low	low	low	low	low
Gullbas ⁵⁷	2016	Low	low	Low	low	low	low	low	low
Junod ⁵⁶	2016	moderate	low	Low	moderate	low	low	low	moderate
Nesbitt ⁵²	2014	low	low	Low	moderate	low	moderate	moderate	moderate
Jackson ⁴⁹	2014	low	low	Low	low	low	Low	Low	Low
Gauthier ⁵⁴	2014	Low	Low	Low	Low	Low	Low	Low	Low
Pelgrim ⁵⁵	2012	Low	Low	Low	Low	moderate	Low	moderate	moderate
Canavan ⁵⁰	2010	Low	Low	Low	moderate	moderate	moderate	moderate	moderate

The number of determinants in each individual study ranged from seven to ten respectively. The five key determinants appeared in seven of the ten studies in this systematic review. It may be worthwhile to consider including these five key determinants in feedback and performance assessments.

Feedback delivery is influenced by a number of factors. One of them, according to research, is that the examiner lacks the ability to translate his observation into detailed, non-judgmental, and constructive feedback^{3,60}. Therefore, feedback will ultimately be ambiguous and meaningless to students seeking to improve their performance⁶⁰.

Effective feedback tools, from the perspective of educators, should include determinants that aid in the learning process, such as helping students comprehend their subject area and providing clear guidance on how to enhance their learning. Structuring feedback by using the five identified determinants will improve alignment between GRS and observed marks. That will lead to a better understanding GRS in observed marks.

Developing a digital tool to evaluate written feedback from OSCE will help in cases where there is a discrepancy between observed marks and the Global Rating Scale result. In these cases, written feedback can be utilized in case of a pass/fail decision. This could demonstrate the significance of feedback in decision-making, as well as how written feedback is viewed as a learning tool that leads to improvements in student performance. The five identified determinants with the highest kappa values could be used as a method to quantify written feedback. Tutors and educators should be made aware of these determinants prior to the OSCE so that they provide beneficial feedback to students. Having a structured comments section could also help overcome the writing challenges tutors currently face when marking the OSCE.

Further study is needed to categorize determinants and sub classify them as part of a quantification approach. This digital measurement tool in medical education will help improve students' performance and knowledge acquisition.

This systematic review had some limitations. For example, grey literature was not included in the study and we reviewed only English studies which may mean results are not generalizable. Another limitation is the focus on written feedback in one type of clinical skills assessment (OSCE). Future research should also consider other training feedback in postgraduate training as well as undergraduate training.

Conclusion

This work suggests that good quality written feedback should be specific, balanced, and constructive in nature, and should describe the gap in student learning as well as observed behavioural actions in the exams. Integrating these five core determinants in OSCE assessment will help guide and support educators in providing effective and actionable feedback for the learner.

Data availability

All data underlying the results are available as part of the article and no additional source data are required.

Reporting guidelines

Zenodo: PRISMA checklist and flow diagram for 'A systematic review of effective quality feedback measurement tools used in clinical skills assessment. <https://doi.org/10.5281/zenodo.6003404>⁶¹

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgments

The authors would like to express their gratitude to Rosie Dunne for her assistance in performing this systematic review by explaining how to conduct a scientific search strategy using key relevant words.

The authors acknowledge Moher *et al.*, (2009) for the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement.

Authors' contributions

AA: took the lead in writing the manuscript and played a key role in systematic literature review. DL: was involved in editing the figures and tables as well as was one of the reviewers in marking the studies. MN: provided a significant input in interpretation of the result as well as manuscript editing. TK: was involved in reviewing, editing and supervising the work. All authors read and approved the final version of the manuscript.

Notes on contributors

AA, (MBBS, MSc), is a PhD student at the School of Medicine, National University of Ireland, Galway. His research interest is with assessment in medical education. He is a teaching assistant in medical education department, in the college of medicine, of Taif University at Saudi Arabia as well. ORCID: <http://orcid.org/0000-0002-0965-5834>

DL, is a Bachelor student at the School of Medicine, National University of Ireland, Galway. Her research interest is with assessment in medical education.

MN, (BA, HDip, MSc, PhD), is the academic director of innovative postgraduate programme in Medical Science and Surgery at National University of Ireland, Galway. His research interests include: exercise rehabilitation for cancer, physiological performance and assessment in sports and exercise, and lifestyle medicine. ORCID: <http://orcid.org/0000-0001-7920-0811>

TK, (BSc PT, PPT, MA, MSc, PhD), is a Senior Lecturer in Medical Informatics and Medical Education at the National University of Ireland, Galway. His research interests include postgraduate medical education and continuing professional development. ORCID: <http://orcid.org/0000-0001-7071-3266>

References

1. Wallace J, Rao R, Haslam R: **Simulated patients and objective structured clinical examinations: Review of their use in medical education.** *Adv Psychiatr Treat.* 2002; **8**(5): 342–348.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Carr S: **The Foundation Programme assessment tools: An opportunity to enhance feedback to trainees?** *Postgrad Med J.* 2006; **82**(971): 576–579.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Cantillon P, Sargeant J: **Giving feedback in clinical settings.** *BMJ.* 2008; **337**(7681): a1961.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Rowe A: **The personal dimension in teaching: why students value feedback.** *International journal of educational management.* 2011; **25**(4): 343–360.
[Publisher Full Text](#)
5. Schartel SA: **Giving feedback - an integral part of education.** *Best Pract Res Clin Anaesthesiol.* 2012; **26**(1): 77–87.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Patrício MF, Julião M, Fareira F, et al.: **Is the OSCE a feasible tool to assess competencies in undergraduate medical education?** *Med Teach.* 2013; **35**(6): 503–514.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Rush S, Ooms A, Marks-Maran D, et al.: **Students' perceptions of practice assessment in the skills laboratory: An evaluation study of OSCAs with immediate feedback.** *Nurse Educ Pract.* 2014; **14**(6): 627–634.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Bouchoucha S, Wikander L, Wilkin C: **Nurse academics perceptions of the efficacy of the OSCE tool.** *Collegian.* 2013; **20**(2): 95–100.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Selim AA, Ramadan FH, El-Gueneidy MM, et al.: **Using Objective Structured Clinical Examination (OSCE) in undergraduate psychiatric nursing education: Is it reliable and valid?** *Nurse Educ Today.* 2012; **32**(3): 283–288.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. McWilliam PL, Botwinski CA: **Identifying strengths and weaknesses in the utilization of objective structured clinical examination (OSCE) in a nursing program.** *Nurs Educ Perspect.* 2012; **33**(1): 35–39.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Gledhill RF, Capatos D: **Factors affecting the reliability of an objective structured clinical examination (OSCE) test in neurology.** *S Afr Med J.* 1985; **67**(12): 463–467.
[PubMed Abstract](#)
12. Cunningham JP, Neville AJ, Norman GR: **The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE.** *Adv Health Sci Educ Theory Pract.* 1996; **1**(3): 227–233.
[PubMed Abstract](#)
13. Brown G, Manogue M, Martin M: **The validity and reliability of an OSCE in dentistry.** *Eur J Dent Educ.* 1999; **3**(3): 117–125.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Hodges B: **Validity and the OSCE.** *Med Teach.* 2003; **25**(3): 250–254.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Setyonugroho W, Kennedy KM, Kropmans TJ: **Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review.** *Patient Educ Couns.* 2015; **98**(12): 1482–1491.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Regehr G, MacRae H, Reznick RK, et al.: **Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination.** *Acad Med.* 1998; **73**(9): 993–997.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Rajiah K, Veettil SK, Kumar S: **Standard setting in OSCEs: A borderline approach.** *Clin Teach.* 2014; **11**(7): 551–556.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Pell G, Homer M, Fuller R: **Investigating disparity between global grades and checklist scores in OSCEs.** *Med Teach.* 2015; **37**(12): 1106–1113.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Abraham RM, Singaram VS: **Using deliberate practice framework to assess the quality of feedback in undergraduate clinical skills training.** *BMC Med Educ.* 2019; **19**(1): 105.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Ryan T, Henderson M, Ryan K, et al.: **Designing learner-centred text-based feedback: a rapid review and qualitative synthesis.** *Assess Eval High Educ.* 2021; **46**(6): 894–912.
[Publisher Full Text](#)
21. Archer JC: **State of the science in health professional education: Effective feedback.** *Med Educ.* 2010; **44**(1): 101–108.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Al-Mously N, Nabil NM, Al-Babtain SA, et al.: **Undergraduate medical students' perceptions on the quality of feedback received during clinical rotations.** *Med Teach.* 2014; **36** Suppl 1: S17–S23.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Wardman MJ, Yorke VC, Hallam JL: **Evaluation of a multi-methods approach to the collection and dissemination of feedback on OSCE performance in dental education.** *Eur J Dent Educ.* 2018; **22**(2): e203–e211.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Griffiths JM, Luhanga U, McEwen LA, et al.: **Promoting high-quality feedback: Tool for reviewing feedback given to learners by teachers.** *Can Fam Physician.* 2016; **62**(7): 600–602 and e419–e421.
[PubMed Abstract](#) | [Free Full Text](#)
25. McMahon-Parkes K, Chapman L, James J: **The views of patients, mentors and adult field nursing students on patients' participation in student nurse assessment in practice.** *Nurse Educ Pract.* 2016; **16**(1): 202–208.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Chapman CL, Burlingame GM, Gleave R, et al.: **Clinical prediction in group psychotherapy.** *Psychother Res.* 2012; **22**(6): 673–681.
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Moher D, Liberati A, Tetzlaff J, et al.: **Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement.** *BMJ.* 2009; **339**(7716): b2535.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Wray CD, McDaniel SS, Saneto RP, et al.: **Is postresective intraoperative electrocorticography predictive of seizure outcomes in children?** *J Neurosurg Pediatr.* 2012; **9**(5): 546–551.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. McHugh ML: **Interrater reliability: The kappa statistic.** *Biochem Med (Zagreb).* 2012; **22**(3): 276–282.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Sterne JA, Hernán MA, Reeves BC, et al.: **ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions.** *BMJ.* 2016; **355**: i4919.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Schünemann HJ, Cuello C, Akl EA, et al.: **GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence.** *J Clin Epidemiol.* 2019; **111**: 105–114.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Zhang S, Soreide KK, Kelling SE, et al.: **Quality assurance processes for standardized patient programs.** *Curr Pharm Teach Learn.* 2018; **10**(4): 523–528.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Schartel SA: **Giving feedback - An integral part of education.** *Best Pract Res Clin Anaesthesiol.* 2012; **26**(1): 77–87.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Norcini J, Burch V: **Workplace-based assessment as an educational tool: AMEE Guide No. 31.** *Med Teach.* 2007; **29**(9): 855–871.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Aeder L, Altshuler L, Kachur E, et al.: **The "Culture OSCE"—introducing a formative assessment into a postgraduate program.** *Educ Health (Abingdon).* 2007; **20**(1): 11.
[PubMed Abstract](#)
36. Burgess A, Mellis C: **Receiving feedback from peers: Medical students' perceptions.** *Clin Teach.* 2015; **12**(3): 203–207.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Bienstock JL, Katz NT, Cox SM, et al.: **To the point: medical education reviews—providing feedback.** *Am J Obstet Gynecol.* 2007; **196**(6): 508–513.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Bing-You R, Hayes V, Varaklis K, et al.: **Feedback for Learners in Medical Education: What is Known? A Scoping Review.** *Acad Med.* 2017; **92**(9): 1346–1354.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Busari JO, Weggelaar NM, Knottnerus AC, et al.: **How medical residents perceive the quality of supervision provided by attending doctors in the clinical setting.** *Med Educ.* 2005; **39**(7): 696–703.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Halman S, Dudek N, Wood T, et al.: **Direct Observation of Clinical Skills Feedback Scale: Development and Validity Evidence.** *Teach Learn Med.* 2016; **28**(4): 385–394.
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Solheim E, Plathe HS, Eide H: **Nursing students' evaluation of a new feedback and reflection tool for use in high-fidelity simulation - Formative assessment of clinical skills. A descriptive quantitative research design.** *Nurse Educ Pract.* 2017; **27**: 114–120.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Ingram JR, Anderson EJ, Pugsley L: **Difficulty giving feedback on underperformance undermines the educational value of multi-source feedback.** *Med Teach.* 2013; **35**(10): 838–846.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Schlair S, Dyche L, Milan F, et al.: **A faculty development program to prepare instructors to observe and provide effective feedback on clinical skills to internal medicine residents.** *J Gen Intern Med.* 2011; **26**: S577.
44. Mitchell JD, Ku C, Diachun CAB, et al.: **Enhancing Feedback on Professionalism and Communication Skills in Anesthesia Residency Programs.** *Anesth Analg.*

- 2017; 125(2): 620-631.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Camarata T, Slieman TA: **Improving Student Feedback Quality: A Simple Model Using Peer Review and Feedback Rubrics.** *J Med Educ Curric Dev.* 2020; 7: 2382120520936604.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Tekian A, Borhani M, Tilton S, et al.: **What do quantitative ratings and qualitative comments tell us about general surgery residents' progress toward independent practice? Evidence from a 5-year longitudinal cohort.** *Am J Surg.* 2019; 217(2): 288-295.
[PubMed Abstract](#) | [Publisher Full Text](#)
 47. Beck Dallaghan GL, Higgins J, Reinhardt A: **Feedback Quality Using an Observation Form.** *J Med Educ Curric Dev.* 2018; 5: 2382120518777768.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 48. Karim AS, Sternbach JM, Bender EM, et al.: **Quality of Operative Performance Feedback Given to Thoracic Surgery Residents Using an App-Based System.** *J Surg Educ.* 2017; 74(6): e81-e87.
[PubMed Abstract](#) | [Publisher Full Text](#)
 49. Jackson JL, Kay C, Jackson WC, et al.: **The Quality of Written Feedback by Attendings of Internal Medicine Residents.** *J Gen Intern Med.* 2015; 30(7): 973-978.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 50. Canavan C, Holtman MC, Richmond M, et al.: **The quality of written comments on professional behaviors in a developmental multisource feedback program.** *Acad Med.* 2010; 85(10 SUPPL): S106-S109.
[PubMed Abstract](#) | [Publisher Full Text](#)
 51. Page M, Gardner J, Booth J: **Validating written feedback in clinical formative assessment.** *Assess Eval High Educ.* 2020; 45(5): 697-713.
[Publisher Full Text](#)
 52. Nesbitt A, Pitcher A, James L, et al.: **Written feedback on supervised learning events.** *Clin Teach.* 2014; 11(4): 279-283.
[PubMed Abstract](#) | [Publisher Full Text](#)
 53. Tomiak A, Braund H, Egan R, et al.: **Exploring How the New Entrustable Professional Activity Assessment Tools Affect the Quality of Feedback Given to Medical Oncology Residents.** *J Cancer Educ.* 2020; 35(1): 165-177.
[PubMed Abstract](#) | [Publisher Full Text](#)
 54. Gauthier S, Cavalcanti R, Goguen J, et al.: **Deliberate practice as a framework for evaluating feedback in residency training.** *Med Teach.* 2015; 37(6): 551-557.
[PubMed Abstract](#) | [Publisher Full Text](#)
 55. Pelgrim EA, Kramer AW, Mookink HG, et al.: **Quality of written narrative feedback and reflection in a modified mini-clinical evaluation exercise: An observational study.** *BMC Med Educ.* 2012; 12(1): 97.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 56. Perron NJ, Louis-Simonet M, Cerutti B, et al.: **The quality of feedback during formative OSCEs depends on the tutors' profile.** *BMC Med Educ.* 2016; 16(1): 293.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 57. Gulbas L, Guerin W, Ryder HF: **Does what we write matter? Determining the features of high- and low-quality summative written comments of students on the internal medicine clerkship using pile-sort and consensus analysis: A mixed-methods study.** *BMC Med Educ.* 2016; 16: 145.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 58. Morgan RL, Thayer KA, Santesso N, et al.: **A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE.** *Environ Int.* 2019; 122: 168-184.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 59. Spickard A 3rd, Corbett EC Jr, Schorling JB: **Improving residents' teaching skills and attitudes toward teaching.** *J Gen Intern Med.* 1996; 11(8): 475-480.
[PubMed Abstract](#) | [Publisher Full Text](#)
 60. Brukner H, Altkorn DL, Cook S, et al.: **Giving effective feedback to medical students: A workshop for faculty and house staff.** *Med Teach.* 1999; 21(2): 161-165.
[PubMed Abstract](#) | [Publisher Full Text](#)
 61. Akram A, Xin LDL, Micheál N, et al.: **A systematic review of effective quality feedback measurement tools used in clinical skills assessment.** 2022.
<http://www.doi.org/10.5281/zenodo.6003404>

Open Peer Review

Current Peer Review Status: ? ✓

Version 1

Reviewer Report 08 June 2023

<https://doi.org/10.21956/mep.20297.r33428>

© 2023 Ramani S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ **Subha Ramani** 

Brigham and Women's Hospital, Harvard Medical School, Boston, USA

The authors have provided a comprehensive description of how they conducted their systematic review on the topic of written feedback on OSCE assessments. They make a strong case for improving the quality of written feedback and also using tools to evaluate the quality of these narratives. I agree that faculty training around writing high quality and impactful narratives would be much more useful to learners than simple ratings. Specificity of narratives would guide learners in self-reflection and formulating performance improvement plans. Appropriate methodology and data collection for the purpose of the study, followed guidelines for systematic review and reaching their conclusions.

Only minor comments:

- use of words like insightful feedback will raise the question- insightful as perceived by whom and to whom. Perhaps actionable is more appropriate.
- 'measurement' of the quality of written feedback. The implication is that only numbers can be 'objective'. This seems contradictory when we are talking about written feedback being more meaningful than just numerical ratings
- Lastly, I would like to have seen a sentence on what this study adds to the field. Many studies on verbal feedback have included all these criteria of quality- specific, constructive, actionable etc. Just a sentence that these have been stated for verbal feedback conversations and equally important when clinical teachers are required to provide written or narrative feedback where they are not having conversations with learners.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Qualitative research, Feedback, Educational theories, Mentorship**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 10 Jun 2023

Akram Alsaahafi

Dear Dr Subha Ramani

Thank you for your comments on our manuscript. Your constructive suggestions have significantly enriched our study. We are working on a revised version and will incorporate your valuable feedback. We appreciate your time, expertise, and the significant contribution you have made to our work.

Kind regards,
Akram

Competing Interests: No competing interests were disclosed.

Reviewer Report 05 June 2023

<https://doi.org/10.21956/mep.20297.r33223>

© 2023 Abraham R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Reina Abraham**

Nelson Mandela School of Medicine, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa

Thank you for requesting me to review this paper: A systematic review of effective quality feedback measurement tools used in clinical skills assessment

This was an interesting article evaluating studies that have measured the quality of written feedback in clinical exams to identify determinants for providing quality written feedback. This is significant because developing effective feedback quality measurement tools that include these determinants will improve the quality of written feedback provided by educators, thereby assisting

students in their learning process to improve their clinical skills performance.

The introduction builds a logical case and context for the problem statement. The problem statement is well articulated and the rationale for, and objectives of, the Systematic Review is clearly stated. The literature review is up to date, well integrated and critically appraised.

The study design is stated and is suitable for a systematic review. The search strategy using the PRISMA guidelines is described. Please provide a rationale for why you are limiting your search from Jan 2010? Why not search all? The publication sources and data extraction process are sufficiently described and referenced. Measures to ensure quality of the publication sources are explained.

The statistical analysis is quite robust, and its interpretation are reported correctly and appropriately.

The search results are organized and easy to understand. The determinants of quality feedback is clearly indicated and appropriate. However, the way the reviewers came about scoring each of the 14 studies against the list of determinants in Table 2, as one +, two +, three +, four + or one -ve, two -ve etc is not explicitly described. Two tables and a figure are presented and agree with the text. I would suggest rephrasing the text for figure 1: Flow diagram showing study selection process based on Preferred Reporting Items for Systematic Reviews and Meta- Analyses guidelines²⁷.

The discussion is in the same order as the result findings. The feedback quality determinants are correctly interpreted and compared with the literature. The conclusion is drawn adequately and is supported by the results presented in the systematic review. The study limitations are explained. Practical significance or theoretical implications are discussed and guidance for future studies is offered.

Title and abstract - The title is representative of the content and breadth of the study. The title captures the importance of the study and the attention of the reader. The abstract is complete. The results in the abstract are presented in sufficient detail. The conclusion in the abstract is justified by the information in the abstract and the text. No inconsistencies between the abstract and the text.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Jun 2023

Akram Alsaifi

Dear Dr Reina Abraham

We are grateful for your insightful feedback and recommendations on our manuscript "A Systematic Review of Effective Quality Feedback Measurement Tools Used in Clinical Skills Assessment". Your comments have been instrumental in helping us improve our work.

In response to your query regarding our decision to limit our search to papers from January 2010 onward, our primary aim was to capture the most recent trends and developments in written feedback for clinical skills assessments. We chose this timeframe based on preliminary research indicating a significant surge in publications related to our topic of interest around 2010. Moreover, considering the resource constraints of conducting this review, we aimed to balance achieving a comprehensive search and managing our resources. We acknowledge that this rationale was not clearly stated in our paper, and we will amend the methodology section accordingly.

Regarding scoring studies against the determinants listed in Table 2, our approach involved a panel of four reviewers who independently evaluated each study. A determinant was assigned a plus score (+) if it was explicitly addressed within the study, with multiple plus scores (++) or (+++) for a stronger presence of a determinant. Conversely, an unaddressed or absent determinant was assigned a minus score (-), with multiple minus scores (-- or ---) if the absence was particularly noteworthy. This led to each study receiving an aggregate score. We will enhance the transparency of this scoring methodology in our revised manuscript.

Lastly, to improve the clarity of the caption for Figure 1, we propose revising it to: "Flow diagram illustrating the process of study selection employed in this systematic review, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines."

Thank you again for your valuable input. We look forward to any further comments or suggestions you may have.

Kind regards,
Akram

Competing Interests: No competing interests were disclosed.