Leveling the playing field: Improving oral exam fairness and scoring reliability through the co-creation of an assessment rubric for the oral exam in the FSV Bachelor course in Global Health

Lena Skovgaard Andersen

Section of Global Health, Department of Public Health
University of Copenhagen

Introduction

Oral exams as an assessment method has a long history in Denmark and continues to be a primary form of assessment at universities (Hurford, 2020). There are certain benefits to conducting oral exams including providing students the opportunity to 1) practice their oral presentation skills, often needed in professional and societal contexts (Hurford, 2020), 2) obtain clarifications and guidance on the questions during the exam, and 3) demonstrate the level of their knowledge and understanding of the subject matter by being able to elaborate on their responses (Slavin, 2014). Oral exams also reduce opportunities for plagiarism, which is particularly relevant at the present time with universal access to chatbot artificial intelligence systems such as ChatGPT (King & chatGPT, 2023).

There are, however, also disadvantages to the oral exam assessment method. Oral exams can obviously not be anonymized and therefore can allow for examiner biases, conscious or unconscious, to creep in (Andersen et al., 2019). In addition, reliability of the scores between examiners can be compromised if the assessment criteria are not stipulated and clarified ahead of time. Oral exams can also heighten nervousness and anxiety in students with the potential to impair the presentation of their responses during the exam (Hurford, 2020).

One approach to address the above-mentioned disadvantages and ensure that oral exams are "criteria-aligned, fair and inclusive", which is in line with the DUT guide on oral examinations in Denmark (Hurford, 2000), is to create and implement an assessment rubric for scoring the

oral exam. Since the exam needs to be congruent with the intended learning outcomes (ILOs), the assessment rubric also needs to align with the ILOs. The assessment rubric should specify the assessment criteria and what level needs to be met to obtain scores, which in Denmark are scores ranging from -3 to 12. An assessment rubric helps to promote alignment between the censors when assigning scores during the oral exam, hence improving reliability. Transparent assessment criteria also benefits students by emphasizing the intended learning objectives and clarifying the standard of quality expected in the exam, while also allowing for students to make reliable assessments of their own performance that can promote continuing improvement. Preparing students for what to expect during the exam can also help reduce anxiety.

There is evidence to support the acceptability, utility, and effectiveness of the scoring rubric (Reddy, 2010). Recommendations for how to create assessment rubrics have also been published. These include recruiting colleagues to collaborate on developing the assessment criteria and stipulating the standards to be met to accomplish each score. Also, students should ideally be included in the co-creation process to ensure the process is inclusive and fair (Race, 2001), and to promote their understanding of the ILOs and its relationship to the course assessment.

The Bachelor degree in Public Health (i.e. FSV) at the University of Copenhagen includes a course in Global Health (GH). The purpose of the course is to "provide the student with knowledge, tools and skills to analyze and understand health problems specific to low- and middle-income countries" (https://kurser.ku.dk/course/sfob20011u). The course is currently organized around themes (for example, there is a unit on mental health that lasts 5 weeks). Each theme is connected to a geographical region (e.g. Sri Lanka) where our lectures have a long history of research collaboration. Each lesson is three hours long. The first two hours are lecture based. During the last hour students work on two questions in groups. Slightly amended versions of these questions are then included in the pot of potential questions that can be picked by chance during the summative oral exam. The only assessment for the course is the summative oral exam. There is currently no assessment rubric for use in the oral exam. To date this has not seemed to be a problem. I have been the censor for the past two years since this exam

format was implemented. Fortunately, the examiner and I have had strong alignment when grading. However, I will not always be the censor and for the sake of continuity and to ensure continued reliability, I believe an assessment rubric would be helpful. The aim of the current project was therefore to co-create an assessment rubric for the oral exam in the FSV course in Global Health.

Method

The assessment rubric was co-created with lecturers, students, and the course leader through an iterative process.
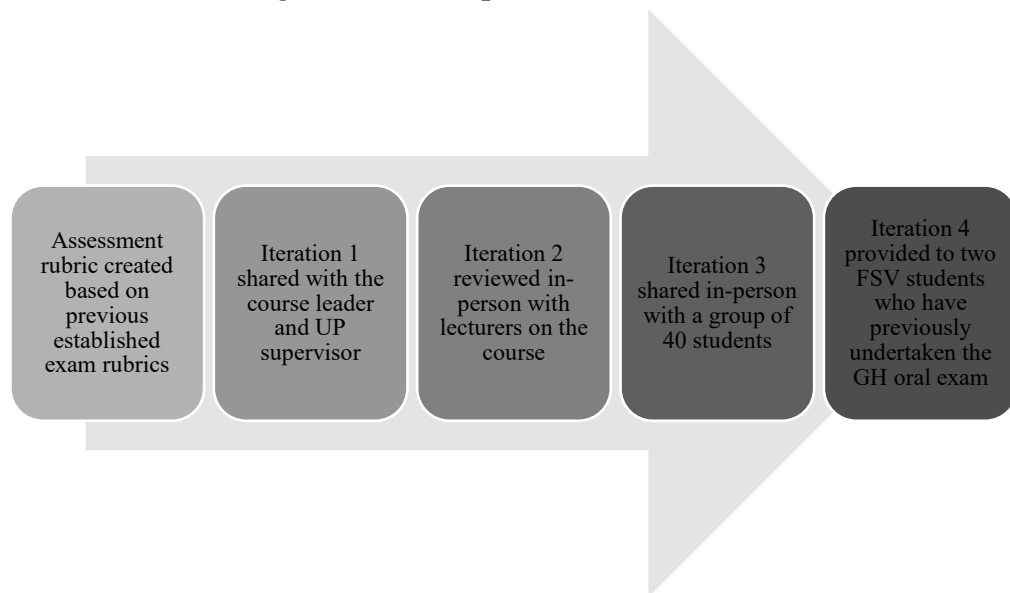
Fig. 1. The iterative co-creation process

1) An initial draft of the assessment rubric was drafted, in part based on a grey literature search, and shared with the course leader and with my local UP supervisor to ensure the assessment rubric aligned with the ILOs of the course and to obtain their preliminary feedback.
2) Then all of the lecturers on the course were invited to an in-person meeting to discuss the course knowledge, skills, and competencies and ensure that they are consistent across lectures and are represented in the intended learning outcomes (ILOs), to ensure that the oral exam aligns with the ILOs, and to provide input on the design of the assessment rubric for the oral exam. The initial draft of the assessment

rubric was circulated ahead of time to provide lecturers the opportunity to review it in preparation for the in-person meeting.

3) A revised version of the oral assessment rubric was then shared with a class of 40 students who were going to be partaking in an oral exam to obtain their feedback and suggestions for improvement of the assessment rubric.

4) As a final step, a revised version of the assessment rubric was shared with two former students of the FSV course who have previously taken part in the Global Health oral exam. They were consulted due to their personal experience with the oral exam process, including the specific exam questions, and with the specific ILOs of this course. This final step also served a 'quality control' function

Results

The first iteration of the oral exam assessment rubric based on the combination of previously established assessment rubrics can be seen in Appendix 2. Seven assessment criteria were included: Comprehension, Argument, Evidence, Presentation, Structure, Prompting, and Ability to answer questions. Each assessment criteria had a predefined weight and were scored across the five standards of Great, Good, Decent, Needs developing, and Unacceptable.

The course leader was supportive of the assessment rubric for the oral exam and he provided some preliminary feedback on the draft, specifically to remove the assessment criteria relating to the presentation (i.e. maintaining eye contact, animated voice, etc.) as this does not reflect the ILOs. My UP supervisor recommended that the weighted system of each assessment criteria be removed and replaced with standards to be met to accomplish the scores 12 to -3, in line with the Danish grading system, which meant the inclusion of two additional categories of scores. He also suggested to keep the language of each assessment criteria consistent with that used in the KU grading guidelines (chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://healthsciences.ku.dk/about/qualityeducation/quality_assurance/files/Guidelines_for_assessment_and_grading.pdf).

Seven colleagues involved in the course attended the in-person meeting and provided input on the assessment criteria and the standards to be met to accomplish each score. The main suggestions included to combine certain assessment criteria to eliminate overlap. Specifically, they recommended combining the argument and structure criteria and adding the prompting criteria and the ability to answer questions to the comprehension criteria. Consequently at this stage in the process, three primary assessment criteria emerged: 1.) Comprehension, 2.) Evidence, and 3.) Argument & Structure. My colleagues also supported the suggestion by the UP supervisor to follow the format and language of the 7-point KU grading scale. The main challenge identified for the ongoing development of the assessment rubric was to be able to clearly distinguish qualitatively between the standards that need to be met for each score, e.g. to obtain a 12 over a 10 or a 10 over a 7, etc.

The student feedback supported the updated assessment criteria and the format of the scoring on the assessment rubric (12 to -3). However, they felt that changes were needed in some places regarding the level of performance across the score descriptors. Specifically, they felt that the 10 for understanding should have a slightly better performance compared to the 7 and that for a 2 the performance should be better than what was stipulated. See Table1 below recommendations for iteration 3.

Table 1. Feedback provided by 40 students

Category	Good (7)	Fair (4)	Inadequate (0)
Comprehension	Shows a good understand-ing of the subject. Knows the material but does not substantially is unable to connect to-pics within or across lect-ures to support argument-ation.	Demonstrates a moder-ate understanding of the subject. Knows some of the material and makes corr-ect claims, but cannot ad-equately connect to-pics.	Shows rudimentary no under-standing and does not cannot for-mulate relevant answers
Argument / Structure			Does not to any significant degree formulate a position or argument.
Evidence	Supports argument with scientific/professional refe-rences that are mostly rel-evant and/or mostly accu-rate. Presents sufficient or most-ly sufficient professional re-ferences to support argume-ntttttttation	Supports argument with scholarly/profess-sional references, that are mostly relevant and /or mostly accurate, but missing important ref-erences. Presents limited evide-nce to support argum-ent	No or Vvery little or irrelevant supsup-port of argument. Presents scientific/profe-ssional or non-scientific refe-rences that are inac-curate and/or irelevant.

Unfortunately, I was unable to obtain feedback from the two former FSV students and hence complete the final iteration as planned. The final version of the oral exam assessment rubric after the previous iterations can be seen in Appendix 1.

Discussion

Through this iterative, co-creation process, a more refined assessment rubric was developed. Since an assessment rubric should be an ever evolving tool that is consistently improved upon (Reading, 2018), my intention is to use this assessment rubric as the basis for discussion with the FSV students this coming semester. Further improvements will be made while providing transparency and clarity to the students and giving them an opportunity to take ownership over their learning. There are a number of ways to engage the students with the oral exam assessment rubric. For example, exemplars can be provided for the students to apply the assessment rubric to or students can provide feedback to each other's work using the rubric (Reading, 2018). Given the current format of the course, my plan is to have the students apply the rubric to the questions that they work on in groups at the end of each lecture, i.e. the questions which form the basis for the oral exam. Students can then explain their grading and scores and we can discuss it as a class. Further improvements can then be made to the assessment rubric.

Particular focus should be placed on the challenging areas identified during this project. First, clearly differentiating each level of performance across the scores can be tricky. Even the language used in the University of Copenhagen 7-point grading scale descriptors is not clearly distinctive, for example, the difference between an 'excellent' (12) performance and a 'very good' (10) performance. Second, since it was recommended in the current project to remove the explicit weighting of each criteria, it is important to ensure that the students and the examiners understand which criteria carry greater importance than others (Reading, 2018). Specifically, the criteria of understanding and evidence may factor more highly in the scoring than argument and structure. It is important to achieve consensus on this at the beginning of the next

semester with the course leader and the students. The outcome must also be communicated to the censor.

A limitation of the current project is that no FSV students have yet to provide input or feedback on the oral exam rubric, so the feedback has been more general for oral exams overall and not specific to the FSV Global Health course. It could, however, be beneficial to have a general oral exam rubric for use across oral exams in the Section of Global Health or the entire Department of Public Health. This more generic assessment rubric could then be tailored through a co-creation process with the students in different courses that use the oral exam format. The current iteration of this oral exam assessment rubric (shown above) could perhaps serve this purpose.

As stipulated in the literature, it would be beneficial to test the validity and the reliability of the assessment rubric. The inter-rated reliability could be assessed by multiple assessors (including the course leader and censor) independently scoring the same oral exam using the assessment rubric. Test-retest reliability could be assessed across several semesters with similar exam questions to determine if the scores remain consistent. It would also be useful to evaluate the construct validity of the rubric and the internal consistency of the criteria.

Conclusion and Personal reflections

With the prominence of oral exams in the Danish university system, it is imperative to implement tools that promote transparency, clarity, and reliability and validity in scoring. The assessment rubric is one such tool. I hope that the current version of the oral exam assessment rubric can form the basis for further co-creation efforts with the FSV students on the Global Health course. I also believe this rubric has the potential for use in other courses and I encourage course leaders to use this assessment rubric as a basis for co-creation efforts in their own courses.

References

Andersen, H.L., Dahl, B. & Tofteskov, J. (2019) Oral exams. In L. Rienecker, P.S. Jørgensen, J. Dolin & G.H. Ingerslev (eds.), *University Teaching and Learning*. Copenhagen: Samfundslitteratur: 373-374.

Hurford, D. (2020). *DUT Guide: Strategies for criteria-aligned, fair and inclusive oral exams*. Dansk Universitetspædagogisk Netværk (DUN), årgang 16, nr. 29.: 119-127. http://dun-net.dk

King, M.R. & chatGPT. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher

Education. *Cel. Mol. Bioeng.* **16**, 1–2 (2023). https://doi-org.ep.fjernadgang.kb.dk/10.1007/s12195-022-00754-8

Race, P. (2001). *A briefing on self, peer and group assessment*. Assessment Series No. 9. LTSN

Generic Centre.

Reddy, Y.M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*:4, 435-448, DOI: 10.1080/02602930902862859

Slavin, R.E. (2014). Cooperative learning and academic achievement: Why does groupwork work? *Anales de psicologia, 30* (3) 785-791.

Appendix

Appendix 1

Category	Excellent (12)	Very Good (10)	Good (7)	Fair (4)	Minimally adequate (2)	Inadequate (0)	Unacceptable (-3)
Comprehension	Shows a deep, robust understanding of the subject. Is able to connect topics from within or across lectures to support argumentation. Demonstrates extensive knowledge of the subject by answering the question accurately	Shows a very good understanding of the subject. Is able to connect topics from within or across lectures to support argumentation, but makes some slight errors. Demonstrates very good knowledge of the subject by answering the questions accurately and appropriately, but requires some degree of probing by the examiner.	Shows a good understanding of the subject. Knows the material, but does not substantially connect topics from within or across lectures to support argumentation. Demonstrates good knowledge of the subject by answering the questions accurately and appropriately, but fails to elaborate.	Shows a fair understanding of the topic. Knows some of the material and makes correct claims, but cannot connect adequately connect topics. Demonstrates some knowledge, but the answers were incomplete or superficial.	Shows a superficial understanding of the subject. Knows some of the material but is unable to connect topics and/or makes false claims. Requires many clarifying and follow-up questions. Responses are unclear (required further follow-up) or incomplete and/or superficial.	Shows rudimentary understanding of the subject and does not formulate relevant answers. Shows incomplete knowledge of the subject by answering questions inaccurately.	No show
Evidence	Supports argumentation with relevant scientific, professional references. Provides sufficient evidence to support argument.	Supports argumentation with scientific, professional references that are mostly relevant and/or mostly accurate. Provides sufficient or mostly sufficient professional references to support argumentation	Supports argumentation with scientific, professional references that are some-what relevant and/or mostly accurate, but lack important references. Provides limited evidence to support argument	Supports argumentation with scientific, professional references that are somewhat relevant and/or mostly accurate, but lack important references. Provides limited evidence to support argument	Limited support for argumentation. Provides scientific/ professional references that are irrelevant.	Very little or irrelevant support for argumentation. Provides scientific/professio nal or non-scientific references that are inaccurate and/or irrelevant.	
Argument and Structure	Clearly articulates a position or argument. Ideas and information are presented in a logical order. Introduction describes the argument well, and establishes a framework for the rest of the response. There is a well-established and articulate conclusion.	Articulates a position or argument, but is at times unfocused. Ideas and information are presented in a logical order with few digressions. Introduction describes the argument well, and establishes a framework for the rest of the response. There is a well-established conclusion with only slight weaknesses in clarity.	A position or an appropriate argument is put forward, but it is not clearly formulated. Ideas and information are presented mainly in a logical order, with a few areas of incoherence or lack of clear progression. Introduction describes the argument well, and establishes a framework for the rest of the response. There is a well-established conclusion, but it is somewhat unclear or unfocused.	Formulates an opinion or argument that is incomplete or limited in scope. There are a few areas of lack of coherence or intermittent lack of logical progression of ideas. Introduction does not lay out the argument well and does not establish a framework for the rest of the answer. There is a conclusion, but it is unclear, illogical, and/or unfocused.	Formulates an opinion or argument that is incomplete, limited in scope, unfocused, or ambiguous. Ideas are somewhat incoherent and/or don't always flow logically, making it a bit hard to follow. Introduction does not lay out the argument well and does not establish a framework for the rest of the answer. Introduction does not lay out the argument well and does not establish a framework for the rest of the answer. There is no conclusion	Does not to any significant degree formulate a position or argument. Ideas are incoherent and/or do not flow logically, which is why reasoning is impossible to follow.	

Appendix 2

Category (weight)	Great	Good	Decent	Needs developing	Unacceptable
Understanding (30%)	Shows a deep/robust understanding of the topic. Is able to connect topics from within or across lectures to support argument.	Shows a good understanding of the topic. Knows the material, but is unable to connect topics from within or across lectures to support argument.	Shows a moderate understanding of the topic. Knows some of the material and makes correct assertions, but is unable to connect topics.	Shows a superficial understanding of the topic. Knows some of the material, but is unable to connect topics and/or makes incorrect assertions.	Shows no understanding of the topic. Unable to answer.
Argument (10%)	Clearly articulates an appropriate position or argument	An appropriate position or argument is made, but is not articulated clearly.	Articulates a position or argument that is incomplete or limited in scope	Articulates a position or argument that is incomplete, limited in scope, unfocused or ambiguous.	Does not articulate a position or argument.
Evidence (30%)	Presents evidence that is relevant and accurate Presents sufficient amount of evidence to support argument	Presents evidence that is mostly relevant and/or mostly accurate Presents sufficient or mostly sufficient evidence to support argument	Presents evidence that is mostly relevant and/or mostly accurate Presents limited evidence to support argument	Presents evidence that is somewhat inaccurate and/or irrelevant, but corrects when prompted Does not present enough evidence to support argument, but	Presents a lot of inaccurate and/or irrelevant evidence Doesn't present enough evidence to support argument, even when prompted repeatedly

				augments when prompted	
Presentation (10%)	Delivery is clear. Uses a clear, audible voice. Some degree of eye contact. Appropriate terminology and language is used. Some degree of animation.				Delivery is unclear. Difficult to hear or understand. No eye contact. Inappropriate terminology and language is used. No animation.
Structure (10%)	Ideas and information are presented in a logical sequence. Introduction lays out the argument well, and establishes a framework for the rest of the answer. There is a well-stated conclusion.	Ideas and information are mainly presented in a logical sequence with a few areas of disjointedness or lack of clear progression. Introduction lays out the argument well, and establishes a framework for the rest of the answer. There is a well-stated conclusion.	There are a few areas of disjointedness or intermittent lack of logical progression of ideas. Introduction does not lay out the argument well and does not establish a framework for the rest of the answer. There is no well-stated conclusion.	Ideas are somewhat disjointed and/or do not always flow logically, making it a bit difficult to follow. Introduction does not lay out the argument well and does not establish a framework for the rest of the answer. There is no well-stated conclusion.	Ideas are disjointed and/or do not flow logically, hence argument is very difficult to follow
Prompting (5%)	Did not have to prompt with probing	Only had to ask one or two probing questions	Had to ask several probing questions.	Required a lot of probing questions.	Required a lot of probing questions.

	questions at all.	and answers were accurate and complete.	Answers were appropriate but incomplete or superficial.	Answers were unclear (requiring further probing) or were incomplete and/or superficial.	Unable to answer questions.
Ability to answer questions (5%)	Demonstrates extensive knowledge of the topic by responding accurately and appropriately to the question and elaborating beyond the question asked.	Demonstrates good knowledge of the topic by responding accurately and appropriately to the questions, but fails to elaborate.	Demonstrates some knowledge of rudimentary questions by responding accurately to questions, but fails to elaborate.	Demonstrates some knowledge of rudimentary questions but responses are superficial and/ or incomplete.	Demonstrates incomplete knowledge of the topic by responding inaccurately and inappropriately to questions.