

Instructions for Analyzing Exam Questions

These instructions help to check the quality of closed-ended questions as part of a post-review of a written examination. By analyzing the exam questions, it is possible to identify which questions (items) may require review, adjustments to the answer key and scoring, removal from the examination or other revisions for further use.

The instructions are aligned with Moodle's Quiz statistic reports (Quiz > Results > Statistics; please also refer to the Moodle documentation¹).

The steps listed below should be followed for all closed-ended questions, because even questions that have an unremarkable statistical result when analyzed may contain ambiguities at the level of response options.

Please note that the interpretability of the statistical key figures depends on the number of students: The larger the number, the more reliable these key figures become.

Step 1: Check the students' comments

If the students had the opportunity to write comments on the questions during the examination, these comments provide invaluable information on possible errors and ambiguities in questions.

Step 2: Check the column «Facility index»

The facility index provides information about the difficulty of the question. It is represented as the average percentage of points achieved in the question by all students.

From a measuring point of view, Krebs (2019)² recommends aiming for a range of 50 - 90 % correct answers for multiple choice items. He recommends considering questions with a facility index of less than 40 % as potentially too difficult.

Considerations

- Does the facility index of the questions correspond to the intended value when the questions were created?
- Are there questions that were answered significantly better or worse than expected?

Interpretation

- A **high** facility index may be due to:
 - Questions that test a core concept that you want students to know.
 - Icebreaker questions that were deliberately placed at the beginning.
 - Poorly constructed incorrect response options (distractors) that reveal the correct answer as the only plausible one.
- A **low** facility index may be due to:
 - Questions about very challenging concepts, that do not match learning objectives, or about concepts that have not been adequately explained.
 - Questions at the end when students were under time pressure.
 - Questions that are unclear or misleadingly phrased, or that contain multiple plausible response options.

Step 3: Check the column «Discrimination index»

The discrimination index is the correlation between the scores of the question and the total scores of the rest of the examination (without the current question).³

The discrimination index should only be interpreted, if the statistics were calculated for a (sub)test that does not contain any open-ended tasks, such as essay-type questions or programming exercises.

Considerations

- Was a question answered correctly by those students with a high total score?

Interpretation

A high value indicates that a question differentiates well between students with varying levels of performance in the remaining examination (without the current question).

A further review of a question is particularly indicated if it has a **low facility index and a low discrimination index**. Following limit values are orientation aids:

> 50%	Very good discrimination
30 - 50%	Adequate discrimination
20 - 29%	Weak discrimination*
0 - 19%	Very weak discrimination*
negative	Question probably invalid*

* Low values in the discrimination index should be interpreted with caution: An individual question may correlate poorly with the overall score, but still be highly relevant in relation to a sub-competency. If the sub-competencies assessed in a course unit correlate only weakly with each other, all highly valid questions for a sub-competency may systematically show low discrimination indices.

Step 4: Check the column «Frequency» of response options¹

The frequency data indicates how often a response option was answered correctly, incorrectly or not at all.

Considerations

- How were the individual response options answered?
- Was an incorrect response option chosen more often than expected **or not at all?**

Interpretation

The frequency data of each response option is useful to help identify poorly designed incorrect response options (e.g. no students selected it) or incorrect response options that are in fact plausibly correct (e.g. question is not clear or response options are not different enough).

Step 5: Identify the cause of a conspicuous question

- Check to make sure the answer key is correct.
 - For single-choice items, the correct response option has to be the unambiguously best answer to the question.
 - In all other cases, the correct response options need to be unambiguously correct and the incorrect response options (distractors) unambiguously incorrect.
- Check if the content was covered in detail so that students can adequately answer the question. For this, you may need to review the teaching materials of the course unit.
- Discuss the question with content experts, co-examiners or students.

Step 6: Make decisions about a conspicuous question

*Decision alternatives for the **current** examination⁴*

- Question remains unchanged in the examination.
- Adjust the answer key and regrade the examination.
- Remove the question from the evaluation of the examination, e.g. by manually assigning a full score to all students.

*Decision alternatives for **further** examinations*

- Question remains unchanged in the question bank.
- Adjust the answer key in the original question bank.
- Modify the content of the question.
- Remove the question from the question bank.

¹ Moodle documentation:

- **Structure and description of the Moodle page:**

https://docs.moodle.org/en/Quiz_statistics_report

The frequency of response options is displayed by clicking on the name of a question in the Moodle page "Statistics".

- **Calculation formulas:** https://docs.moodle.org/dev/Quiz_statistics_calculations

- **A guide** to the statistics from the Open University:

https://docs.moodle.org/403/en/images_en/e/e9/Brief_Guide_to_iCMA_reports.pdf

² Krebs, R. (2019). Prüfen mit Multiple Choice (S. 51). Springer.

³ Discriminative efficiency shown in the graphic is not discussed here. Please refer to the Moodle documentation¹.

⁴ ETH Zurich, Educational Development and Technology. (2013). Guidelines on Grading Written Examinations (1st edition). https://ethz.ch/content/dam/ethz/main/eth-zurich/organisation/let/files_EN/guidelines_grading.pdf