



You are here: ► University of Vienna (<http://www.univie.ac.at/en/>) ► Teaching Affairs and Student Services (<http://studien-service-lehrwesen.univie.ac.at/>) ► Center for Teaching and Learning (<http://ctl.univie.ac.at/>) ► Infopool besser lehren (/en/) ► Start Page (/en/start-page/) ► Assessment & Grading (/en/start-page/assessment-grading/) ► Assessing Exams

Assessing Examinations

Reinhard Schott

November 2018

Overview

1. Reference Standards for Assessment

2. Assessing Written and Oral Exams

2.1. Multiple-choice Questions and Written Exams with Assessing Exams

Simple Open-Ended Questions

2.2. Written Exams (Essay Questions) and Oral Exams

2.3. Grading Key

3. Quality Criteria of Exams

4. Assessment Errors

4.1. Avoiding Assessment Errors

University exam grades fulfil several **functions**: they score specific performances, provide students with feedback, and inform other universities and potential employers of the performance levels of graduates.^[1] Grades that **differentiate between different student performance levels** based on previously defined intended learning outcomes and criteria will fulfil this function.

1. Reference Standards for Assessment

In order to evaluate a performance fairly, it is best to use **previously defined intended student learning outcomes (knowledge and competencies to be acquired by students) as the reference standard**. These learning outcomes should be communicated to students. Furthermore, the **criteria** for “attaining the intended learning outcome” (e.g. “students must attain more than 50% of total points for a positive grade” – see “grading key” below [link]) must be defined. Using **criteria-oriented references** requires absolute **performance standards** (i.e., **to what extent has the examinee achieved the predefined performance standards or intended student learning outcomes?**).^[2]

In **practice** at the university, social context can influence exam situations, especially in the case of oral exams. **Orientating grading on social norms (i.e. an individual's performance is assessed relative to that of peers), however, is problematic in terms of study law regulations**; in addition, it is unfair to students. When exams are evaluated according to social norms, we assume that students' performances follow a standard normal distribution (**Gaussian distribution**). In extreme cases, this leads to the best performers of an exam cohort to receive the best grade (“Sehr gut”), and the worst performers to receive a failing grade (“Nicht genügend”)—regardless of whether or not they have achieved the intended student learning outcomes. Moreover, exam participants from different cohorts may receive different grades for the same performances. **Assessing according to social norms is neither compatible with the Austrian universities act nor with the statutes of the University of Vienna.**

2. Assessing Written and Oral Exams

All methods of assessment are similar in the sense that **model solutions** and/or **criteria for evaluating answers** allow a standardised and fair evaluation of the performances. The most **practical advantage of model solutions** is that individual exams (from different examiners) can be fairly evaluated according to the same schemes and standards. Teachers may also use such model solutions and criteria when discussing grades with students during exam review sessions or in the case of grade disputes (see: Universities Act § 79 (5) and § 84).

The complexity of developing criteria/model solutions varies depending on the kind of exam; In any case, it should be done in advance. **As a general rule**, the greater the freedom that students have in solving problems or answering questions in assignments or exams, the more demanding it is to determine and break down the criteria for evaluating answers. Students may get a better sense of the exam requirements if they are provided example questions with model solutions, as they prepare for an exam.

2.1 Multiple-choice Questions and Written Exams with Simple Open-Ended Question

In these cases, establishing **criteria for evaluating answers** is usually easy, because you may use “correct/incorrect choice or mark” or “correct/incorrect word” as criteria. For example, each correct answer receives one point (or a half-point for a partly correct answer), and if students attain a certain number of points, they will receive a passing grade.

	MC Questions	Short Written Answers
Example criteria for evaluating answers	marked correctly not marked marked partly correctly	correct incorrect partly correct

2.2 Written Exams (Essay Questions) and Oral Exams

We recommend formulating model solutions and/or criteria for evaluating answers in advance, especially when assessing open-ended questions (written and oral). A model solution should include all important aspects and concepts. Model solutions are especially suitable for questions that have one correct answer. According to Bloom’s classification,^[3] these are mainly knowledge, comprehension, and application questions.

An additional possibility is to develop **individual criteria** for each question that, taken together, result in an **assessment scheme**.

Open-Ended Written or Oral Questions

Example criteria for evaluating answers:	all partial aspects were stated; the required examples were provided; the key technical terms were applied correctly; relationships were recognised and demonstrated; solutions were presented and justified; the argumentation is coherent; etc. ...
---	---

Example: A written knowledge-based question with a model solution and criteria (assessment scheme)

Question 1

- What are nociceptors, where do they occur, and how are they different from other sensory receptors? (3 points)

Model solution:

- Nociceptors are pain receptors that are present throughout the entire body except in the brain and lungs. In contrast to other sensory receptors, repeated or continuous pain stimulus does not lead to adaptation but instead to the opposite, i.e. sensitisation.

(Possible) criteria for evaluating answers

- Student explains what nociceptors are (1 point)
- Student indicates where nociceptors are found in the body (1 point)
- Student explains the difference between nociceptors and other sensory receptors (1 point; if only adaptation or sensitization is mentioned, then ½ point)
- Any additional information that was not required but is relevant (1 point)
- Overall, a maximum of 3 points is possible

2.3 Grading Key

Evaluate student performance based on established **evaluation criteria** (grading key). Grades reflect the extent to which students meet the requirements that are derived from the intended student learning outcomes (criteria-oriented reference standard). Use these **criteria** to determine the grade of students rather than the performance of other students. **Generally, teachers determine the number of points that students need to attain a passing grade (pass-fail threshold)**. Teachers also determine the distribution of the remaining grades. **In practice, students are often required to achieve “more than 50%” of possible points to pass.** Between the passing threshold and the maximum number of attainable points, the distribution of the remaining grades may be linear (the space between a “4” and a “3” is the same as between a “3” and a “2” or a “2” and a “1”). Alternatively, you may use a different distribution method (e.g. the space between a “4” and a “3” is larger than the space between a “3” and a “2,” and that space is in turn larger as the space between a “2” and a “1”).

3. Quality Criteria of Exams

Exams should fulfil specific **quality criteria**.^[4] In order to guarantee or improve the quality of an exam, consider the following important questions when designing or conceptualising your exams:

- **Does the exam measure what it is intended to measure, is the exam valid (validity)?** Does your exam cover the scope of the intended student learning outcomes? Do the exam questions measure the knowledge and competencies as defined in the learning outcomes? In order to get answers to these questions, **colleagues may provide feedback on each others' exam questions**. They may scrutinise exam questions to see if they “measure what the exam is intended to measure,” i.e. if questions measure the intended learning outcomes. For this, you ideally consult colleagues who are familiar with the exam material and the learning outcomes. In addition, colleagues can check if questions inadvertently test for language comprehension or attentiveness.
- **How precisely does the exam measure (reliability)?** An exam is reliable when it measures the examined characteristic (knowledge and competencies in a specific discipline) with a high degree of precision. **A higher number of moderately difficult exam questions generally leads to higher reliability**. If the circumstances in practice allow it, **the same questions** (perhaps in a different order) should be **given to parallel groups**.
- **Are the exam results independent of the examiner (objectivity)?** Are the results of an exam independent of who designs, administers, assesses, or interprets it? Are all students evaluated according to the same criteria? In practice, pre-established **model solutions and/or criteria** designed to evaluate answers, make standardised and fair assessment of all student performances possible.

The following factors, among others, play an additional role: the necessary resources (time, budget, material, etc.) relative to the information gained by an exam (quality criterion of **economy**) and the temporal, psychological, and physical burden for students (**appropriateness**).

4. Errors of Assessment

In order to assess written exams fairly and to implement oral exams free of error and unbiased, you should be familiar with potential evaluation tendencies and errors in order to take countermeasures if necessary.^[5]

Expectation effect: Teachers' positive or negative expectations of students may have a self-fulfilling prophecy effect on the assessment process. Expectations regarding grade distribution may also have unintended effects. Teachers may be reluctant to give a “1” five or ten consecutive times, as this would contradict an implicit assumption that grades have to vary.

Projection error and halo effect: As teachers, we may project our own characteristics, views, wishes, or errors (usually unconsciously) onto our students, which can influence assessment. We may draw conclusions from one characteristic to another, which in fact are entirely independent. The perception of a student may “outshine” the assessed performance (halo effect) if, for example, the clothing, language, handwriting, or attractiveness of a student influences the evaluation of his or her performance. This may, for example, cause a teacher to assess a perceived “talent” or personal characteristics instead of the actual performance.

Sequence effect: The sequence of evaluating exams may influence the results. We tend to grade the first works that we evaluate more strictly as the last ones. Previous exam performances can also have an effect on our evaluation. For example, an average exam is often evaluated better when **we read** it immediately after an extremely poor exam, or vice versa. In oral exams, the examiner remembers the first and last performance best; and thus these performances may have a greater influence on the overall assessment.

Strictness and mildness: Strictness errors occur when even “slight flaws” have a disproportionate large influence on evaluations and “good” performances are almost ignored. In contrast, in the case of **mildness errors** “good” performances weigh heavily, whereas “poor” performances hardly impact the evaluation. Your own performance expectations (e.g. of “young and strict” or “old and forgiving” examiners) should not influence the assessment.

Tendency towards extremes: In this case, a distinction is made between “good” and “poor” performances, above all. If the threshold of “good” has been reached, the work receives the best grade; if it is not reached, it receives the worst grade. Grades in the middle of the spectrum are avoided.

Tendency to the middle: In contrast to the tendency towards extremes, there may also be a tendency to avoid explicit extreme assessment. The tendency to the middle occurs primarily when evaluators are unsure about how to assess a performance.

4.1 Avoiding Assessment Errors

Teachers have various options to minimise assessment errors. In the case of written exams, assessment errors such as the halo effect, sequence effects, expectation effects, or strictness and mildness errors can be minimised by **evaluating one exam question at a time across all participants**, instead of the entire exam of one student after the other.

The “overall impression” of a student work should only form at the end, and not after a couple of questions. **Model solutions and/or evaluation schemes** that can help reduce the number of assessment errors in the case of oral exams are also helpful. Feedback from colleagues, e.g. in the case of oral exams before a committee, can contribute to objective assessment.^[6]

Schedule **breaks** when you grade a large number of assignments/exams, or in the case of oral exams with many students. This helps to reduce fluctuations in the strictness of assessments, contrast effects to previous exams, and effects of your own tiredness.

Furthermore, it can help to consider to what extent you are subject to any specific **assessment tendencies** (strictness or mildness; tendency towards extremes, etc.) and to keep these in mind when grading exams. Especially in the case of oral exams, try to ignore any marked characteristics of students that have nothing to do with their exam performance. Any potential **anger** over a student’s performance should not disproportionately affect the assessment.

References

- [1] Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht der Geschäftsstelle mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates. Hamburg 2012. <http://www.wissenschaftsrat.de/download/archiv/2627-12.pdf>.
- [2] Metzger, Christoph, and Charlotte Nüesch. "Fair prüfen. Ein Qualitätsleitfaden für Prüfende an Hochschulen." St. Gallen: Institut für Wirtschaftspädagogik, 2004.; Walzik, Sebastian. "Kompetenzorientiert prüfen. Leistungsbewertung an der Hochschule in Theorie und Praxis." Opladen and Toronto: Verlag Barbara Budrich UTB, 2012; Zumbach, Jörg, and Hermann Astleitner. "Effektives Lehren an der Hochschule." Stuttgart: Kohlhammer, 2016.
- [3] Bloom, Benjamin S., ed. *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim und Basel: Beltz, 1956/1972.
- [4] Kubinger, Klaus D. *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe, 2009.
- [5] Zumbach und Astleitner, *Effektives Lehren an der Hochschule*, [2]; Walzik. *Kompetenzorientiertprüfen*, [2].
- [6] Zumbach and Astleitner, *Effektives Lehren an der Hochschule*, [2].

Recommended citation

Schott, Reinhard: *Assessing Examinations*. Infopool *better teaching*. Center for Teaching and Learning, University of Vienna, November 2018. [<https://infopool.univie.ac.at/en/home-page/assessment-grading/assessing-exams/>]

This work is licensed under a Creative Commons

Attribution-ShareAlike 3.0 Austria License (CC BY-SA 3.0 AT)

For more information please see: <https://creativecommons.org/licenses/by-sa/3.0/at/>



Contact

T (Telephone): +43-1-4277-120 60

infopool@univie.ac.at

University of Vienna | Universitätsring 1 | 1010 Vienna | T (Telephone) +43-1-4277-0

Sitemap (</en/sitemap/>) | Imprint (</en/imprint/>) | Accessibility (<https://www.univie.ac.at/en/about-us/accessibility/>) | Privacy Policy (<https://dsba.univie.ac.at/en/privacy-policy/>) | Printable version