



Evaluating Quiz Questions

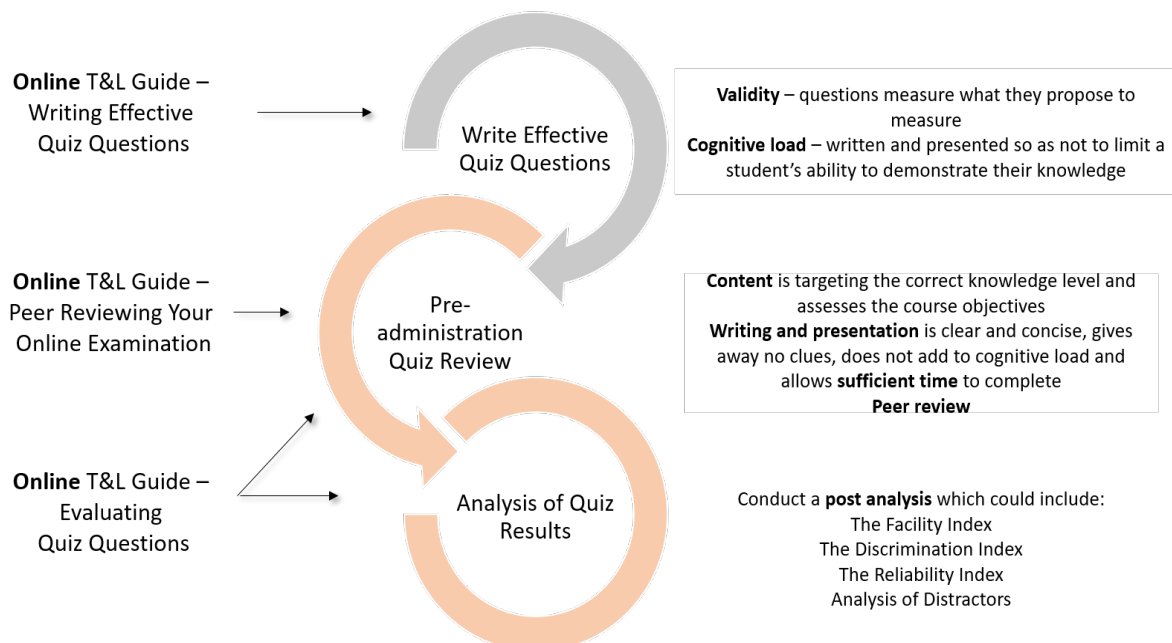
Quizzes are a commonly used online assessment tool – but how do you know if your questions are fair, are able to differentiate strongly-performing students from weaker students and are measuring what you need them to measure?

Writing effective quiz questions can be challenging. Ensuring validity and reliability, as well as controlling for students’ cognitive load, requires thoughtful consideration of what and how we are assessing.

Quizzes provide evidence of students’ development toward learning outcomes. As such, they are data collection instruments, like surveys used for research purposes. Therefore, we must ensure that they are evaluated and refined before they are administered and on an ongoing basis when they are used to assess students’ learning.

In this *Guide*, we discuss quiz evaluation techniques that can be used before a quiz is presented to students as well as what is available after the quiz is completed through the standard quiz reports in learnonline.

Overview





Online quizzes can be evaluated at various times, some of which we have addressed in earlier *Guides* as depicted in the image above. The validity and cognitive load of questions can be evaluated as you write the questions and peer review of the quiz is conducted before the quiz is administered, and a post analysis of the student results happens once the students have completed the quiz.

- Online T&L Guide – [Writing Effective Quiz Questions](#)
- Online T&L Guide – [How do you Peer Review Online Examinations](#)

Pre-Administration Quiz Review

Some of the critical checks that need to be performed before the quiz is administered include:

a) Content

- All questions are designed to assess knowledge at the appropriate level, i.e. they are aligned with verbs in course objectives
- The quiz assesses all relevant course objectives and content areas

A useful tool to enable these checks is a simple matrix that shows content focus and knowledge level for each question included in the quiz. A sample overview matrix for a quiz assessing course objectives at 4 levels of knowledge and 4 topics is provided in the table below.

Knowledge Level	Topic 1	Topic 2	Topic 3	Topic 4	Total items	% of quiz
Comprehend	3	3	3	3	12	20.0%
Apply	4	4	3	4	15	37.5%
Analyse	3	2	3	2	10	25.0%
Evaluate	0	1	1	1	3	7.5%
TOTAL	10	10	10	10	40	100%

b) Question Writing and Presentation

- MCQ Stems are written in clear, accessible language and do not contain extraneous information
- MCQ Keys are the only correct options and are indisputably correct
- MCQ Distractors are plausible

- No 'cuing' is provided in MCQ options
- Questions are presented in ways that capitalise on online affordances without creating unnecessary cognitive load for students (e.g. consider screen layout, use of images, etc.)
- Sufficient time is allocated for completing the quiz. A useful strategy is to time yourself or a colleague from your discipline when completing the quiz and multiplying this time by 3 or 4 to predict how long it might take a competent student. Another rough measurement is to allocate one minute for each simple MCQ, keeping in mind that more complex questions will take longer.

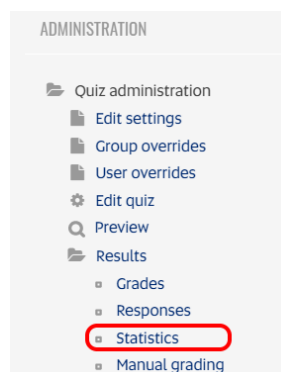
Once you have performed your own checks, it is important to seek feedback on a quiz draft. Colleagues from your discipline can be very helpful as 'critical friends'. If possible, a pilot test should also be conducted by administering the quiz to non-experts, i.e. people who have a similar level of knowledge as your students.

Analysis of Quiz Results

Once the quiz has been administered, either as part of a pilot test or for student assessment, you can use the results to measure the quiz effectiveness. Moodle provides a range of statistical measures in the Quiz Statistics Report that summarises how students interacted with the quiz.

- Moodle – [Quiz Statistics Report](#)
- Resource – [Brief Guide to the Moodle iCMA Reports](#)

The report is accessed from the Administration section (or gear menu) of the quiz itself:

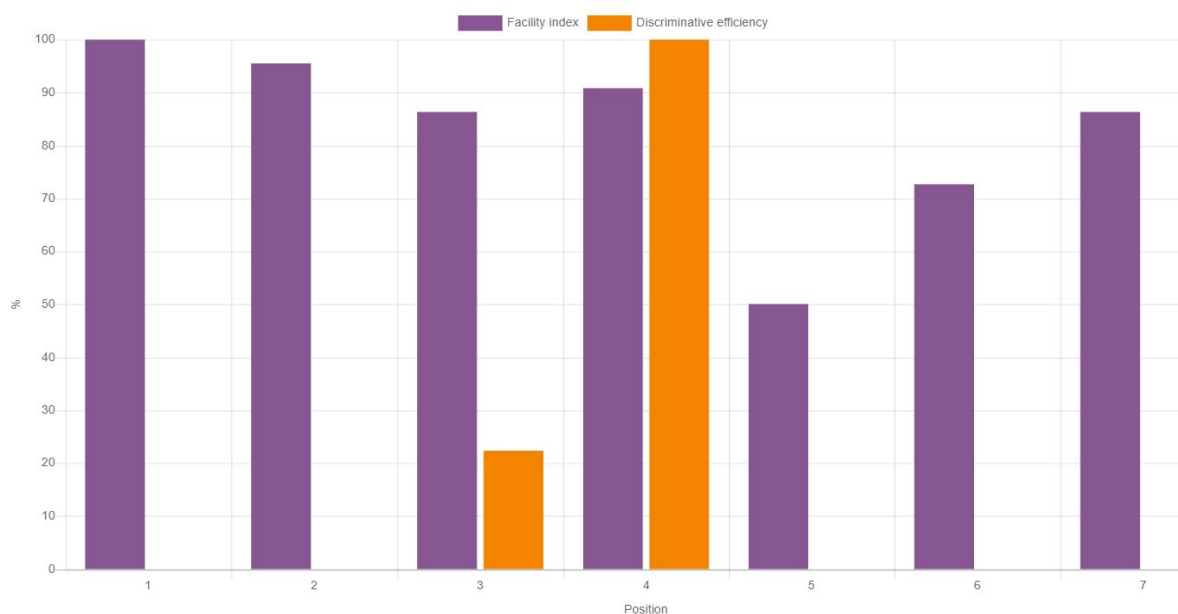


It has several components – a summary of the whole quiz in **Quiz information**, then a **Quiz structure analysis table** providing an analysis of each question (problems highlighted in red) as shown below. Note, when there are random questions, the results for each question in the category are presented.

Q#	Question name	Attempts	Facility index	Standard deviation	Random guess score	Intended weight	Effective weight	Discrimination index	Discriminative efficiency
1	Random [question name]	22	90.91%	29.42%		14.29%	15.00%	33.15%	52.17%
1.1	[question name]-3	5	100.00%	0.00%	25.00%	14.29%			
1.2	[question name]-2	11	90.91%	30.15%	25.00%	14.29%		34.64%	52.17%
1.3	[question name]-1	6	83.33%	40.82%	25.00%	14.29%		90.07%	100.00%
2	Random [question name]	22	86.36%	35.13%		14.29%		-28.20%	-34.97%
2.1	[question name]-2	14	78.57%	42.58%	25.00%	14.29%		-15.15%	-18.07%
2.2	[question name]-1	8	100.00%	0.00%	25.00%	14.29%			

Finally, there is a **Statistics for question positions** graph of two evaluation tools, the Facility index and the Discriminative efficiency index, as shown below for a 7-question quiz.

Statistics for question positions



Investigating Facility Index together with Discrimination efficiency index provides useful data to evaluate each question and can help you to identify questions that need to be modified to provide all students with a fair grade (Gamage et al, 2019).

Facility Index

As the name suggests, the facility index is a measure of how easy a question was for students. It is simply the percentage of students who answered a question correctly.

If less than 5% of the students answered the question correctly, the question is either extremely difficult or there is something wrong with it – for example, it is worded in a way that makes it difficult for students to understand the question.

Conversely, if over 95% of your students have answered it correctly, it is too easy and does not discriminate between successful and unsuccessful learners. However, there may be times when it is appropriate to have a facility score of 95%. An example would be when there is a critical concept that all students must acquire; in all other cases, we should aim for no more than 75% for 4-item MCQs (Parkes & Zimmaro 2016).

If a question records a facility index lower than 30%, it should be reviewed to identify whether it is attached to a concept that has not been appropriately scaffolded in the course or whether there are any issues in the way the question is presented to students.

The Facility index, as well as the other measures discussed in this *Guide*, can also be used to ensure that alternative questions in randomised quizzes are of comparable difficulty. In other words, if one question is randomly selected out of two equivalent items for inclusion in a quiz, then the Facility indexes for the two possible questions should be similar.

Discrimination Efficiency Index

The main idea behind the Discrimination Efficiency Index is that students who perform well on the quiz overall will tend to answer other questions correctly as well. In other words, it is a correlation measure.



A low discrimination efficiency index indicates that students' answers on a specific question are out of pattern – for example, student A answered 10 questions incorrectly but got question X correct, whereas other students who did well overall answered question X incorrectly.

Questions that have a discrimination efficiency index lower than 0.20 or 20% should be discarded (Parkes & Zimmaro 2016).

Internal Consistency

Consistency is measured on the whole quiz as opposed to individual items and is reported in the Quiz information section of the report. The **coefficient of internal consistency** of a quiz measures to what degree individual items correlate with each other and it is a score ranging from 0 to 1. This data is available in the **quiz information** section of the statistics report.

High-stakes, standardised tests should achieve an internal consistency above 0.9. However, for course assessments, an internal consistency index of 0.6 or above is acceptable (Parkes & Zimmaro 2016).

Analysis of Distractors

Ideally, distractors should only be selected by students who have not mastered the content you are assessing. If a distractor is selected by many students, it may indicate that the question or options are not clear. Similarly, if a distractor is never or very rarely selected, then it is probably implausible and should be discarded.

To check whether the distractors are working the way they are intended, we can inspect a frequency table that shows which options were selected by high-performing versus low-performing students. An example of this process is provided in Parkes and Zimmaro (2016).

You can access the detail on which answer options students selected in learnonline by selecting the question from the **Quiz structure analysis table**.

Conclusion

Writing a quiz is only the beginning in a process of continuous improvement. Item and quiz analysis can provide valuable information



not only about the quiz validity and reliability, but also on the effectiveness of the learning process. If an item is excessively difficult, this may indicate that the concept it is assessing has not been appropriately scaffolded. Therefore, a quiz can be a valuable measure of teaching effectiveness, as well as student learning.

If you would like to ask online teaching and learning questions related to your course, you can look through our [FAQs](#), write to TIU@unisa.edu.au, have an [online consultation with a member of the TIU](#) or complete the online modules as part of [Introduction to Engaging Learners Online](#).

References

Gamage, SHPW, Ayres, JR, Behrend, MB & Smith, EJ 2019, '[Optimising Moodle quizzes for online assessments](#)', *International Journal of STEM Education*, vol. 6, no. 1, pp. 1–14.

Parkes J & Zimmaro D 2016, *Learning and assessing with multiple-choice questions in college classrooms*, Routledge.