MACQUARIE University
SYDNEY·AUSTRALIA

QUIZ ANALYSIS

Learning and Teaching Quick Guide

LEARNING AND TEACHING STAFF DEVELOPMENT

Analysing the performance of iLearn (Moodle) quiz questions can help improve the validity of quiz assessment tasks. iLearn provides statistical reports for both the quiz overall and the individual questions.

QUIZ STATISTICS

To access statistical reports from a quiz page, go to the: *Gear icon (actions menu), Results,* then choose *Statistics.*

Use the first attempt to generate the statistics (this is the default). Including multiple or latter attempts may invalidate the statistical results and reccomendations.

The quiz performance report provides indicators as to how well the quiz is discriminating between high and low ability students. Descriptive statistics of overall scores include:

Mean (For discriminating, deferred feedback, tests aim for between 50% and 75%), **Median** (the middle score), **Standard deviation** (12% to 18% is good), **skewness** (-1 to 1 is good) and **kurtosis** (0 to 1 is good). **Coefficient of internal consistency** (CIC is Cronbach's alpha. Higher is better. Above 90% is uncommon, above 75% is satisfactory, below 64% indicates a problem). **Error ratio** (By CIC. Aim for under 50%. See T1). **Standard Error** (aim for 8% or less, greater than 12% indicates a problem).

T1: Error ratio by CIC

CIC	51	64	75	84	91	96	99	100
Error ratio	70	60	50	40	30	20	90	0

QUESTION ANALYSIS

The analysis of each question includes the statistics in T2.

T2: Quiz structure analysis for each question

Attempts: Should equate to the number of students (assuming 1 attempt). Lower values indicate students are not seeing or attempting the question.

Facility index: Average score as percentage. Higher values indicate easy questions. Ideally in the range 30% to 70%. Check questions outside this range (too easy/hard).

Standard deviation: The spread of marks for the question. This will depend on the question difficulty.

Random guess score: Applies to questions with multiple defined responses. This is the expected score if students just guessed. An average of all options. Values above 40% are unsatisfactory.

Intended weight: How much this question was supposed to contribute to determining the overall test score, as per quiz design. Compare it to the next statistic.

Effective weight: How much this question did contribute to the variance in the students' test scores. Effective and intended values should be similar. If not, check for easy, hard or broken questions.

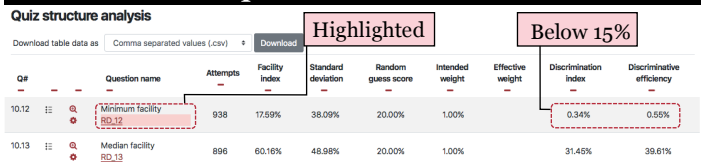
Discrimination index: The correlation of the overall test score and this question score. Students who score high on the test, should also score high on this question. Higher values (30 or above) are better. Low values need attention. Note: discrimination requires a Facility Index in the range 30% - 70% in order to be valid. See also next item.

Discriminative efficiency: An estimate of how good the discrimination index is relative to the difficulty (Facility index) of the question. Low values need attention.

CHECKS

Problematic questions: iLearn will highlight suspect items where the *Discriminative efficiency* is *less than 15%* with a red background (see F1). Problems may include: confusing wording, a repeated question, a previous question provided an answer hint, question is on a topic not taught, or the response options are incorrectly configured (e.g. duplicated or wrong choice set as correct).

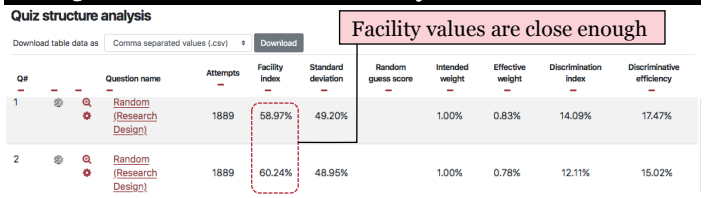
F1: Problematic questions - Discrimination



Too easy or too hard: Check the Facility index for items outside the range 30% to 70%. Consider if appropriate.

Fair random questions from a 'category' or when using **variations of a question**: It is important to ensure each question within the pool is of equal difficulty. Expand a 'random' question group and check that the *Facility index* column values are similar. Values that are too *different* are of concern (say, more than 10. See F2). The less questions displayed in a quiz, the greater is the impact of variation of difficulty in a given pool on the overall fairness of the test.

F2: Question variants - Facility index



MULTI RESPONSE DISTRACTOR EFFICIENCY

Good distractors test for misconceptions, these should be plausible, but wrong choices. Check individual question distractor statistics (via the question name in the report) for choices that are selected by less than 5% of students. Distractors rarely selected by students need attention or may be redundant. Wrong distractors selected by the majority may be incorrectly configured, have unclear wording, not be adequately different, or may it may rightly indicate a common misconception held by students.

Explore this topic further ...

- Using a quiz for online exams https://goto.mq/finalassess
- Butcher, P. (2010) Brief guide to the Moodle iCMA Reports, Open University, UK. Reproduced at https://docs.moodle.org/dev/Quiz_report_statistics
- Moodle docs: Quiz Statistics Calculations https://docs.moodle.org/dev/Quiz_statistics_calculations
- iLearn guides https://goto.mq/ilearn-activities-quick-guides
- Quiz Randomisation Quick Guide and more: https://goto.mq/qg-lt

Acknowledgements: L&T Professional Development Team.
Prepared by Dr Mathew Hillier, Macquarie University, Sep 2022

Find out more about Learning and Teaching at Macquarie University
Macquarie University NSW 2109 Australia
Professional.Learning@mq.edu.au | https://staff.mq.edu.au/teach

ABN 90 952 801 237
CRICOS Provider 00002J

MACQUARIE University