



UNIVERSITÀ DEGLI STUDI
DI GENOVA

**EVALUATION STRATEGIES IN THE MEDICAL SETTING:
Multiple choice testing for critical thinking
and its classroom and remote computer-based protocols
vs. standardized and non standardized oral examination**

DINOGLI SCUOLA DI SCIENZE MEDICHE E FARMACEUTICHE

Prof. Carlo Pesce, MD PhD

GENERAL MEDICAL COUNCIL

The principles of assessment

Schemes of assessment must support the curriculum and allow students to prove that they have achieved the curricular outcomes. This means assessments must allow students to demonstrate the breadth and depth of their knowledge, and to show what they can do. Professional attitudes and behaviour must also be assessed.

Student performance in both the core and elective parts of the curriculum must be assessed and must contribute to their overall result. Students who have not satisfied the examiners in both parts of the curriculum must not be allowed to graduate.

Medical schools should use a range of assessment techniques that are appropriate for testing the curricular outcomes. Medical schools should determine the most appropriate scheme of assessment for their curriculum. However, schemes must meet best practice in assessment, and medical schools must be able to provide evidence that the schemes are valid and reliable, and that they have processes for setting standards and making decisions about student performance.

When students get close to graduating, their knowledge, skills, attitudes and behaviour must be thoroughly assessed to determine their fitness to practise as PRHOs.

FOSTERING CRITICAL THINKING

Critical thinking competence is crucial for professional performance

Students who know how to think make better judgements than those who merely memorize facts. Basic concepts learned in medical school should be applied in a variety of clinical settings. Assessment of educational effectiveness should include a written plan for the evaluation of all the activities, based on outcomes.

Student outcomes must be demonstrated in terms of attained competences, knowledge, skills, and attitudes. These issues largely depend on critical thinking in clinical judgement and decision making.

Critical thinking combines knowledge, context and reflective skepticism. Its acquisition depends on the skills being taught through mastering discipline-related language abilities.

The practice of writing-up and participation in work groups are helpful, eg, through games such as,

What if? What else? What then?

These encourage participants to connect events and make data-based decisions.

CRITICAL THINKING EVALUATION

What are the tools to measure critical thinking abilities and evaluate the effectiveness of teaching this skill?

Closed-end, multiple-choice test questions are difficult to write, but easy to administer. With currently available scoring and item-analysis softwares, they are also easy to analyze.

In medicine, knowing should imply the ability to apply knowledge to clinical scenarios.

Test writing is requires some training but in principle is not difficult to learn. It should be included in the faculty educations programs.

ORAL EXAMINATION

An endangered species?

“It seems important to start with the forthright acknowledgment that no single assessment method can provide all the data required for judgment of anything so complex as the delivery of professional services by a successful physician.”

Miller GE. The assessment of clinical skills/ competence/ performance. Acad. Med. 65: S63-67. 1990

ORAL EXAM

Oral assessment should be used to examine only those capabilities that are best measured by it, and should not be the sole means of examination.

STANDARDIZED vs. NON_STANDARDIZED ORAL EXAM

Content of examinations should be standardized as far as possible.

VALIDITY/RELIABILITY/FAIRNESS

Examining authorities should establish the **validity**, **reliability** and **fairness** of their examinations and publish the pertinent data.

VALIDITY

Validity concerns the capacity of assessment to prompt or allow a candidate to demonstrate the extent to which they possess the requisite knowledge, skills and values.

RELIABILITY

Reliability means “repeatability” or “consistency”. The same result / a similar result should be obtained at different sessions. Assessment is unreliable if a candidate is likely to get a different result with different examiners, with a different sub-set of questions, at a different time, or in a different context.

The oral examination is commonly used to assess clinical knowledge and skills in both undergraduate and postgraduate medical education. In an oral examination, the trainee interacts with the examiner and is assessed based on answers provided to the questions asked.

the interview style, in which the examinee is quizzed on general topics;

the clinical style, in which questions regard specifically diagnosis and treatment plans for a particular patient;

the cognitive style that requires problem solving around specific cases;

the role-playing style, with students assuming various “roles” with the examiner.

FAIRNESS/POTENTIAL BIAS

Fairness implies that candidates of equal standing should on average earn the same test score, irrespective of group membership. It entails both the absence of bias within the test and assessment processes. Examiners should give all candidates an equal opportunity. Bias in tests refers to construct-irrelevant components that result in systematically lower or higher scores for identifiable groups of examinees.

Because of the potential of oral assessment for bias, and the resulting discrimination in sub-groups of candidates, perceptions of bias should be addressed formally and professionally in the context of best examination practice.

STRONG POINTS

TRADITIONAL ORAL EXAMINATION EXAMINATION

The ability of assessing students' reasoning skills

"Oral exams allow you to assess their (students') reasoning and communication skills, and to converse with them."

"I believe the oral exams prepare students for clinical reasoning needed to practice medicine."

Demonstrates students' learning

"Gives students an opportunity to show their skills in a different and more "applicable" way."

"They (students) present their own cases and are usually knowledgeable about the case."

STANDARDIZED ORAL

Uniformity

"Consistency in general clinical knowledge assessment."

"Uniform question list for representative cases that allows more direct comparison of examinees' knowledge base"

Fairness

"Keeps the examiner grounded as to the level of knowledge and reasoning that should be expected for a medical student. Fairness."

"Improves examiner organization and consistency. Standardization attempts to eliminate unconscious bias and allows there to be a system of fair and equitable evaluation."

Easy to use

"Helps less experienced faculty examiners administer the oral exam more easily."

WEAK POINTS

TRADITIONAL ORAL EXAMINATION

No grading standards

“Somewhat subjective”

“Grading sometimes seems to vary among examiners”

“Lack of standardization”

Strong individual variations

“Variation in examiner skills, biases and techniques.”

“The variability in each student’s case list and presentation.”

Time pressure

“I think it takes a lot of time, which is something most of us have little of.”

“Takes students away from the wards an additional 2-3 hours per block”

STANDARDIZED ORAL EXAMINATION

Inflexible

“The standardized questions aren’t always appropriate for the case the student has listed.”

“At times the students’ choice of patient does not quite fit into the category of questions so I end up not completely sticking to the standardized questions.”

“The flow of the oral exam as a “conversation” is made more challenging by the standardized questions -- this is a minor weakness.”

The ability of assessing students’ higher order reasoning skills by having to focus on standardized questions

“Focusing on standardized questions left less opportunity to assess (students’) higher-order clinical reasoning.”

WRITTEN TESTS

Proper, critical-thinking oriented items should

(1) include their rationale

(2) be clearly presented

(3) require multilogical thinking

(4) require ability to discriminate between plausible alternatives.

Testing serves the purpose of evaluation but is also a learning tool.
Self analysis of one's answering process is *per-se* educational.

Test questions should present not only the information indicating the correct answer, but also why distracter are incorrect.

Time spent in careful preparation of critical-thinking oriented questions saves Faculty time in the classroom.

Test question discussion in the class are not useful to promote individual critical thinking.

Test bank and analysis systems (eg ParSYSSTEM) generate reports that help reviewing question quality.

Each student's performance is evaluated and scored.

Critical-thinking tests should reach a higher cognitive level than simple comprehension, that is at the application level or above according to Bloom's taxonomy.

TABLE 1
Verbs Associated with Categories of Cognition for Bloom's Taxonomy of Educational Objectives

Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Define	Describe	Apply	Analyze	Compose	Appraise
Identify	Differentiate	Calculate	Categorize	Construct	Assess
Know	Discuss	Classify	Compare	Create	Evaluate
List	Explain	Develop	Contrast	Design	Judge
Name	Rephrase	Examine	Distinguish	Formulate	
Recognize	Restate	Solve	Determine	Modify	
State	Reword	Use	Investigate	Plan	

MULTILOGICAL THINKING

Multilogical thinking is the ability to view problems from different perspectives. In other words it is thinking that requires knowledge of multiple facts to apply concepts in a logical and systematic fashion in a clinical problem.

If the answer is just a paragraph written in the textbook, the test question does not require critical thinking.

Which assessment finding is the most important to define clinical care protocol for a patient with chronic obstructive pulmonary disease (COPD)?



What are the findings in a patient with COPD?

In a critical-thinking oriented test, the student must learn how to discriminate between plausible alternatives. All possibilities should be viable, but only one is the best choice.

The answer is in the textbook

If the patient says that he can win any war because he is Napoleon, what type of behavior is most important to demonstrate?

To answer this question, the student must know how to deal with delusional patients. The best fitting behavior should be chosen based on concepts of therapeutic communication.

DISTRACTERS

**Writing distracters is the most difficult part of test writing.
An excellent source of distracters is previous discussions
of clinical settings with students.**

QUESTION 1

Which finding is characteristic of a patient with Parkinson's disease?

- A Night blindness
- B Pain in lower extremities
- C Shuffling gait**
- D Incontinence



QUESTION 2

The doctor makes a house call for a 75-y-o patient with a 5-yr history of Parkinson's disease. Which finding has the greatest implication for home care?

- A His grandchildren haven't visited him for over a month
- B There are many throw rugs in the living room**
- C There is a towel wrapped around his neck to wipe his face
- D The sitting patient is gripping the arms of the chair

Question 2

- (1) (1) contains a rationale (absent in question 1)
- (2) (2) is an application level test
- (3) (3) requires multilogical thinking as well as discriminative answering

Critical thinking cannot be examined pre and post entrance into the program, but should be measured systematically as an ongoing process in the clinical context.

As much as the students' medical knowledge increases, so does their ability to apply critical thinking skills to a variety of clinical problems of growing complexity.

USMLE TEST EXAMPLES

A 55-year-old man has had crushing substernal chest pain on exertion over the past 6 weeks. He had a myocardial infarction 2 months ago. He takes nitroglycerin as needed and one aspirin daily. He has smoked two packs of cigarettes daily for 30 years. Examination shows normal heart sounds and no carotid or femoral bruits. Treatment with a β -adrenergic blocking agent is most likely to improve his symptoms due to which of the following mechanisms?

- (A) Decreasing myocardial contractility**
- (B) Dilating the coronary arteries**
- (C) Peripheral vasodilation**
- (D) Preventing fibrin and platelet plugs**

A 28-year-old man comes to the office because of a 1-year history of pain with urination that has increased in severity during the past month. He also has had episodes of blood in his urine during the past 5 years. He lived in sub-Saharan Africa until he came to the USA 6 months ago for graduate school. Temperature is 38.0°C (100.4°F), pulse is 80/min, respirations are 16/min, and blood pressure is 110/84 mm Hg. Physical examination shows suprapubic tenderness.

Laboratory studies shown in table.

Imaging studies show bilateral hydroureter and hydronephrosis and foci of calcification in the region of the bladder. A biopsy specimen of the bladder shows marked chronic inflammation with fibrosis and scattered granulomas. Which of the following best explains the biopsy findings?

- (A) Exposure to a chemical toxin
- (B) Interstitial cystitis
- (C) Malacoplakia
- (D) Schistosomiasis
- (E) Vesicoureteral reflux

Hemoglobin	12.3 g/dL
Hematocrit	37%
Leukocyte count	13,400/mm ³
Segmented neutrophils	65%
Bands	5%
Eosinophils	5%
Lymphocytes	22%
Monocytes	3%
Serum	
Urea nitrogen	75 mg/dL
Creatinine	3.8 mg/dL
Urine	
Blood	3+
RBC	200/hpf
WBC	100/hpf
RBC casts	absent
WBC casts	absent

A 10-year-old boy is brought for a follow-up examination 2 days after he was seen in the emergency department because of hives, hoarseness, and light-headedness. His symptoms began 15 minutes after he was stung by a bee and lasted approximately 60 minutes; they resolved before he was treated. He has been stung by bees three times over the past year, and each reaction has been more severe. Examination shows no abnormalities. Which of the following is the most appropriate recommendation to prevent future morbidity and mortality from this condition?

- (A) Avoid areas known to have bees**
- (B) Avoid wearing colorful clothing outside**
- (C) Carrying diphenhydramine tablets**
- (D) Carrying self-injectable epinephrine**
- (E) Seek immediate medical attention following any future sting**

COMMUNICATION SKILLS

A 26-year-old woman comes to the physician with her husband for counseling prior to conception. Her mother and three of her five siblings have type 2 diabetes mellitus. She is 170 cm (5 ft 7 in) tall and weighs 82 kg (180 lb); BMI is 28 kg/m². Her blood pressure is 148/84 mm Hg. Physical examination shows no other abnormalities. Her fasting serum glucose concentration is 110 mg/dL.

Which of the following is the most appropriate initial statement by the physician?

- (A) "Let's review ways you can optimize your own health before conceiving."**
- (B) "We should test you for islet cell antibodies before you try to conceive."**
- (C) "You can conceive right away since you are in good health."**
- (D) "You should avoid gaining weight during pregnancy because you are already overweight and at risk for type 2 diabetes mellitus."**
- (E) "You should have no problems with your pregnancy if you start insulin therapy."**

TESTING IN THE COMPUTER LAB

The UNIGE Experience



Online exams in the computer lab started in 2013 at UNIGE on a Moodle platform. The same server is being used for access tests through different databases.

The nursing school, with over 1,200 students at six different locations in the Liguria Region, is the largest users. For clinical nursing, run-of-the-mill tests are preceded by critical thinking evaluation sessions including a “triple jump” approach. This is a combined test starting with a clinical case the candidate should study with free access to textbooks. Subsequently, there is a “nursing” test and finally specific tests on single disciplines.

Candidates receive immediate confirmation of each test and their final evaluation. Most tests only offer a deferred feedback for performance revision. The English test, however, shows for each question an immediate feedback on the correct answer. This helps the candidate to manage the time allotted to terminate the test.

The number of examinees varies considerably for each session. The current record has been 400 tests performed in a single day in the 30-seat computer lab. This performance can be achieved using a well-organized team staffing the computer lab and an adjoining classroom. The examinees are summoned in time groups, identified in the classroom and sent to the lab in the “one out one in” fashion.

A computer scientists inside the lab monitors smoothness of the process and results, and checks for possible cheating between examinees sitting at the same desk.

Inglese 1

Questo test di **Inglese 1** è composto di **30 domande** a cui si deve rispondere nel tempo massimo di **35 minuti**. Si supera il test se si risponde correttamente ad **almeno 20 domande**.

Quando hai deciso la risposta che ritieni corretta, puoi cliccare sul pulsante **Check** per confermare. Se vuoi aspettare a dare la risposta puoi scegliere il numero di un'altra domanda da **Quiz navigation** (in alto a sinistra).

Per concludere la prova clicca sul pulsante **Submit all and finish** (in basso).

Fai molta attenzione a rispondere a tutte le domande nel tempo assegnato.

Attempts allowed: 1

This quiz opened at Wednesday, 23 August 2017, 8:00 AM

This quiz will close at Wednesday, 23 August 2017, 2:00 PM

Time limit: 35 mins

Attempt quiz now

QUIZ NAVIGATION

SUFFICIENZA: 20 (SU 30)



Finish attempt ...

Time left 0:29:45

Clinical cases and additional quiz material can be presented in free navigation or in interactive fashion with a blocked sequence with a maximum timing defined by the teacher.

Almost all questions are multiple choice with just one correct—or best—answer. Only rarely the multiple correct answers system is used. If the teacher is experienced in question writing, not only notions but also critical thinking for problem solving can be explored. In comparison with the true/false system multiple choice reduces guessing. However for some disciplines, such as foreign language, the embedded questions with brief answers is preferred.

Security measures include the following: test access allowed only through the IP addressed of the lab, record of time and date of start, high security level of the browser, one question per page, mandatory stop upon completion of test/time allowance.

MATRIX REPRESENTATION OF RISKS IN EXAMS PERFORMED IN THE DIGITAL CLASSROOM



- R1 Fake identity
- R2 Exam performed from elsewhere
- R3 Students exchange tests in classroom
- R4 Two students cheating
- R5 Student access concealed info
- R6 Breach of credentials
- R7 Disloyal invigilating personnel
- R8 Unreliable system support

 From Arrmano et al.
 Atti del MoodleMoot Italia 2018

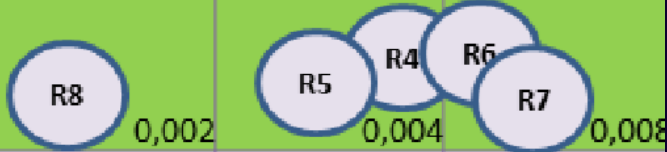
RISK ABATEMENT MEASURES AND THEIR EFFECT

Impact \ Probability	0,05	0,1	0,2	0,4	0,8
0,9	0,045	0,09	0,18	0,36	0,72
0,7	0,035	0,07	0,14	0,28	0,56
0,5	0,025	0,05	0,1	0,2	0,4
0,3	0,015	0,03	0,06	0,12	0,24
0,1	0,005	0,01	0,02	0,04	0,08
0,01	0,0005	0,001	0,002	0,004	0,008

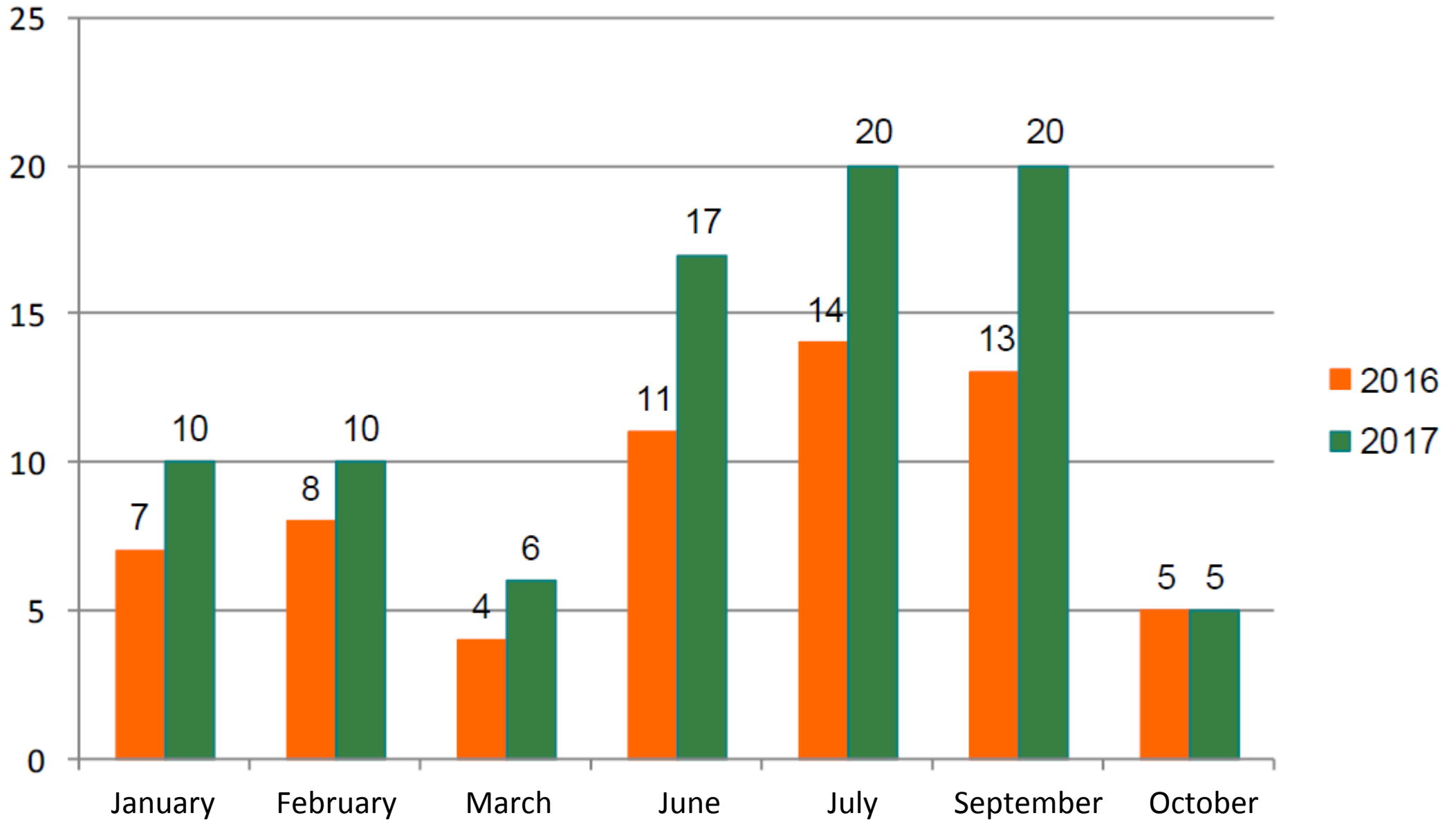
MEASURES

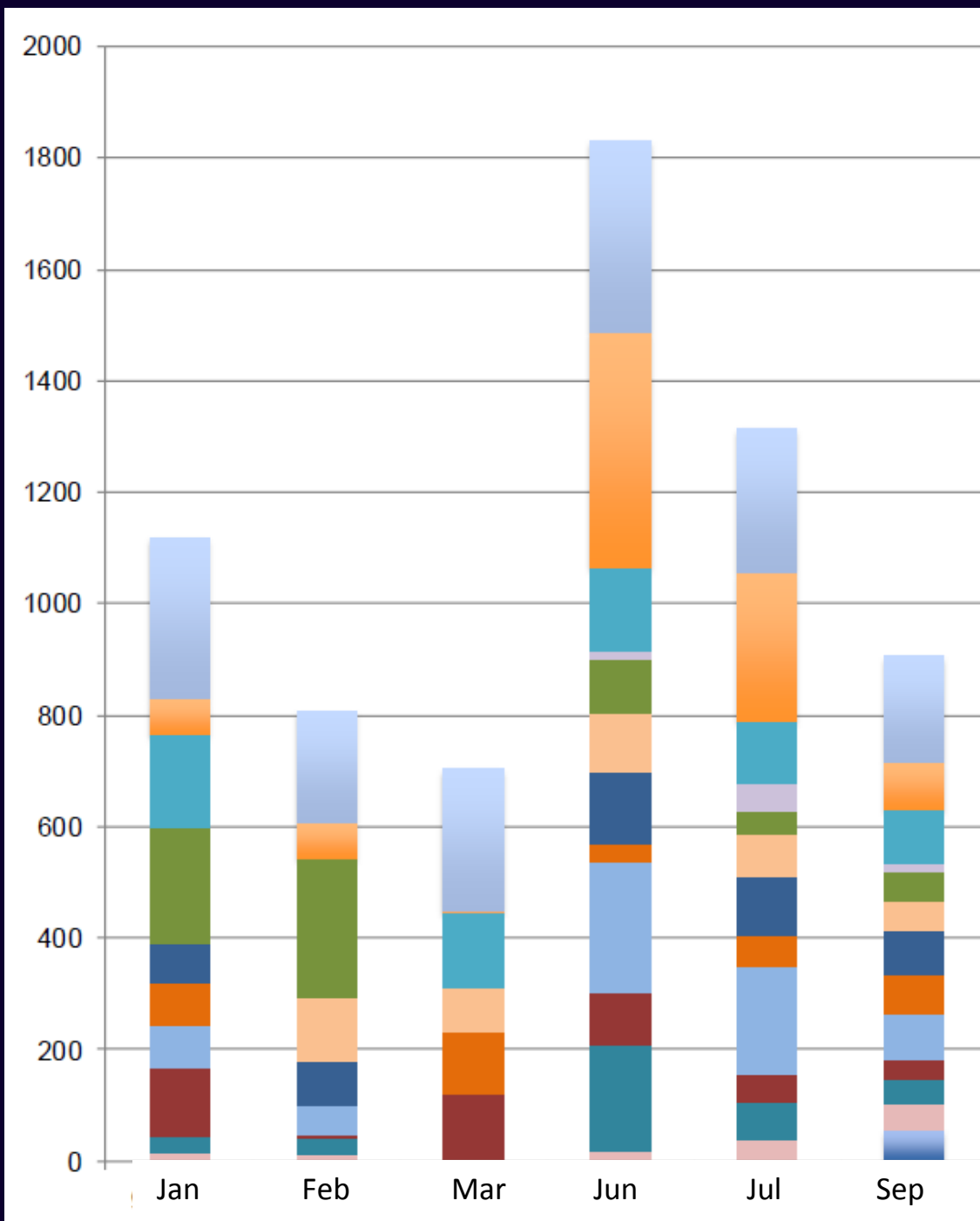
- Personal account access with student's name printed
- Single access permitted
- Identity identification
- Classroom surveillance
- Automated activation of each workstation
- Moodle-dedicated exam software
- Restricted online access

Eradicated risks



EXAM SESSIONS IN DIGITAL CLASSROOM





Number of students sitting for an exam
in the digital classroom by subject

The Quiz is a very powerful activity that can meet many teaching needs, from simple, multiple-choice knowledge tests to complex, self-assessment tasks with detailed feedback.

Questions are created and stored separately in a Question bank and can be reused in different quizzes.

When creating a Quiz you can either make the questions first and add them to the Quiz, or add a Quiz activity (as below) and create the questions as you go along.

STANDARD QUESTION TYPES

Calculated simple

Formula calculation

Calculated multi-choice

As multiple-choice with embedded formula calculation

Drag and drop into text

Students select missing words or phrases and add them to text by dragging boxes to the correct location.

Drag and drop markers

Students drop markers onto a selected area on a background image without predefined areas.

Drag and drop onto image

Students make selections by dragging text, images or both to predefined boxes on a background image.

Description

Info label, no question itself

Essay

This allows students to write at length on a particular subject and must be manually graded.

It is possible for a teacher to create a template to scaffold the student's answer in order to give them extra support.

Matching

A list of sub-questions is provided, along with a list of answers. The respondent must "match" the correct answers with each question.

Embedded Answers (Cloze Test / Gap Fill)

These very flexible questions consist of a passage of text (in Moodle format) that has various answers embedded within it, including multiple choice, short answers and numerical answers.

Multiple choice

With the multiple-choice question type you can create single-answer and multiple-answer questions, include pictures, sound or other media and weight individual answers.

Short Answer

In response to a question, the respondent types a word or phrase. There may several possible correct answers, with different grades.

Numerical

A numerical question allows an acceptable error.

Random short-answer matching

From the student perspective, this looks just like a Matching question. The difference is that the sub-questions are drawn randomly from Short Answer questions in the current category.

Select missing words

Students select a missing word or phrase from a dropdown menu.

True/False

In response to a question, the respondent selects from two options: True or False.

Random question type

A random question is not a question type. It is a way of inserting a randomly-chosen question from a specified category into a quiz.

This means that different students are likely to get a different selection of questions, and when a quiz allows multiple attempts then each attempt is likely to contain a new selection of questions.

Managing your questions

Questions are stored in the course Question Bank






Questions can be organized into Question categories and shared with other teachers.

Moodle has several different formats that can be used to import or export questions.

QUIZ REPORTS

In the course you are logged in, click on your quiz and go to *Administration > Quiz administration > Results*
This will open a submenu that offers reports on:

- Grades report
- Responses report
- Statistics report
- Manual grading

	First name / Surname	Email address	State	Started on	Completed	Time taken	Grade/10.00	Q. 1 /0.83	Q. 2 /0.83	Q. 3 /0.83	Q. 4 /0.83	Q. 5 /0.83	Q. 6 /0.83	Q. 7 /0.83	Q. 8 /0.83
<input type="checkbox"/>	 Sally Student Review attempt	student2@example.com	In progress	10 October 2018 3:19 AM	-	-	-	✓ 0.83	-	✗ 0.00	-	-	-	✓ 0.83	-
<input type="checkbox"/>	 Carla Hague Review attempt	CarlaHague@example.com	Finished	7 January 2021 8:36 PM	7 January 2021 8:37 PM	59 secs	6.14	✓ 0.83	✗ 0.00	✓ 0.83	☑ 0.56	✗ 0.00	✓ 0.83	✗ 0.00	✓ 0.83
<input type="checkbox"/>	 Fannie Asif Review attempt	FannieAsif@example.com	Finished	7 January 2021 8:32 PM	7 January 2021 8:34 PM	1 min 24 secs	4.33	✗ -	☑ 0.33	✓ 0.83	✗ 0.00	✗ 0.00	✓ 0.83	✗ 0.00	✓ 0.83
<input type="checkbox"/>	 Sam Student Review attempt	student@example.com	Finished	9 October 2018 11:00 PM	9 October 2018 11:03 PM	2 mins 58 secs	6.11	✓ 0.83	✗ 0.00	✗ 0.00	☑ 0.28	✗ 0.00	✗ 0.00	✓ 0.83	✓ 0.83
<input type="checkbox"/>	 Rachael Zachary Review attempt	RachaelZachary@example.com	Finished	7 January 2021 8:34 PM	7 January 2021 8:36 PM	1 min 17 secs	5.56	✓ 0.83	✗ 0.00	✗ 0.00	☑ 0.56	✗ 0.00	✗ 0.00	✓ 0.83	✓ 0.83
	Overall average						5.53 (4)	0.67 (5)	0.08 (4)	0.33 (5)	0.35 (4)	0.00 (4)	0.42 (4)	0.50 (5)	0.83 (4)

REFERENCE LINK

A comprehensive course in digital testing on the moodle platform

<https://www.eur.nl/en/course/mooc-assessment-higher-education>



Quiz: automatic grading



Moodle Basics

Question Types

