Blended Learning in Practice

Autumn 2022

Table of Contents

Psychometric analyses of the Canvas Quizzes and STEM Staff's Anxiety

Muhammad Jamro m.jamro@herts.ac.uk

Abstract

Quizzes are the most frequent tools used by academic tutors in the module assessments of the programme of studies. This article focuses on the use of statistical parameters of the Canvas quizzes that possess very intriguing features for academics and learning teaching technologists. Some useful psychometric analyses of the quiz items are scrutinised based on auto-generated statistical parameters. For example, standard deviation; distractors; difficulty, and discriminatory indices, respectively, are effectively used to ensure the quality of questions and overall quizzes for graded assessments. The main thrust of this article is to value the staff's good practices and dedication in developing quizzes fit-for-the purpose despite their increasing workload and encourage young learning technologists and quiz-enthusiasts to invest more efforts in making use of the statistical parameters in evaluating the quality of the quizzes, items in the quizzes, the sustained learning of the students, and fulfilling academic quality assurances' requirements to cover overall learning objectives of courses and programme of studies.

Introduction

Psychometrics is a field of study within psychology and education concerned with the theory and technique of testing, measurement, assessment, and feedback-related activities. This has been even more researched during the ongoing COVID pandemic due to the unprecedented shift of education onto the online domain (DeCoito & Estaiteyeh, 2022; Sim et al., 2021; Telles-Langdon, 2020; Tuah & Naing, 2021). The quizzes are the prominent method of students' assessments when using "virtual" learning management systems (LMSs), which have become dominant online platforms for the delivery of modules [1] and programmes of studies for diverse disciplines ranging from science, technology, engineering, and mathematics (STEM) to non-STEM (or applied-STEM) disciplines. There is healthy literature available for enthusiastic academics and pedagogic critics to compare various features of worldwide dominated learning platforms, e.g., Canvas, Blackboard, Brightspace, and Moodle (IBL_News, 2019; Kiran, 2020; Öztürk & Gürler, 2020; Paynter & Barnes, 2021; Swerzenski, 2021). Looking at the current market trends in recent years, Canvas appears as a new implementation sweeping UK, Ireland, and North American education market sectors (Justin, 2021, 2022). Regardless of their different brands, all these platforms have useful tools and third-party apps for effective delivery of courses for various programmes of

[1] Some universities refer to these as 'courses', both terms are used interchangeably.

studies. Nevertheless, both students and staff have no choice to pick or choose these platforms but to learn about the various features of learning management systems' resources provided by their institutions, whose top hierarchical management decides the choice of platform considering strategic and sustainable solutions. All this increases staff's workload if the learning platform keeps evolving and in a worst-case scenario if there is a switch-over or shift in learning and teaching resources.

The quiz statistics is one of the most interesting, yet not much shared as, good-academic practice. Some staff do publish quiz-related pedagogic research outcomes seeking academic inspiration, however; a majority of academic staff remains detached from potential widening participation, but Schools value staff's professional contributions towards the TEF, REF and KEF frameworks, and departmental portfolios (Johnson, 2022; van Miegroet et al., 2019). Historically, online quizzes became powerful tools for learning, teaching, and assessments (LT&A) with the advent of computer-assisted assessment "CAA" (Bull & McKenna, 2003) and massive open online courses (MOOCs), (Bali, 2014; Shah, 2019; Yıldırım, 2020). A holistic approach is needed in strategic planning to keep the fresh blood pumping in particular the training and development of EdTech professionals and academics supporting the LT&A delivery platform of the institution.

In this article, some intriguing statistical features of Canvas quizzes are shared to help the technology enthusiastic young community of learning technologists, academic staff of STEM, and applied-STEM disciplines to make use of auto-generated statistical data of the quizzes. By adopting these, the staff can help themselves without feeling over-burdened and extend much-needed support to the academic quality assurance (AQA) office, the students, and the academic community, respectively. There are various missing features in current Canvas quizzes that can be implemented if the learning technologists, STEM, and applied-STEM community give their due input, anticipating the future challenges to LT&A and feedback systems. This is a much-needed strategy to achieve greater pedagogic sustainability meeting the requirements of the future education, moving from digital to deep learning and AI-based tools, and industrial learning experience platform (LXP) producing big-data analytics. The interested Edtech members can find more relevant information in the useful list of references (A Kaklij et al., 2020; Atif et al., 2021; Diaz & Young, 2021; Selwyn, 2019).

Quizzes, and Challenges of Modern Learning Management Systems

Historically, the University of Hertfordshire (UH) had its' legacy digital online managed learning environment (MLE) system known as StudyNet for about two decades, which also supported various quizzes features including QuestionMark perception, and tutors practised various online and in-class quizzes for individuals and groups (Jamro, 2017) and electronic voting systems (EVS) 'Clickers' for quizzes and seeking quick feedback via Microsoft® PowerPoint slides based on TurningPoint Technologies® (Jefferies et al., 2013). The former

School of Engineering & Technology[2] immensely used this electronic system issuing the EVS devices to individuals for their entire degree programmes. Staff benefited as individuals in their academic personal development, and programme teams by the virtue of the clicker's statistical software, which produced spreadsheets containing valuable post-assessment data for individuals' and cohort performance analysis.

In 2019, TurningTechnolgies® was acquired by echo360, and recently they have added a smartphone assessment app, replacing the bulky clicker devices, supporting the real-time, in-class, and/or anywhere for both Android® and IOS® operating systems. These apps can be synchronised with Canvas via Learning Tools Interoperability (LTI), which is a new initiative to standardise the integration of third-party or open-source tools with different LMS (IMS-Global, 2019). Currently, there are more than 450 apps, out of which 29 are quizzes related and can be linked to Canvas, but each requires institutional purchase (e.g., LockDown Browser, BigBlueButton, Ouriginal, WebPA, QuestionMark Perception, etc are commonly used by most academics). The interested learning technology developers, instructors, and academic enthusiasts can find more details of various Canvas apps in the web-link [3]. Additionally, Canvas is planning to launch "New Quizzes" in June 2024 with advanced features (Canvas_Community, 2021).

All aforementioned technology enhancements demonstrate that popular software and favourite tools keep evolving and staff adopt them as a regular part of their job. It is noteworthy to mention that recently, with the help of their professional learning technologists, trained staff known as 'StudyNet Champions' and 'Students Technology Mentors', the University of Hertfordshire has completed migration to Canvas making it fit-for-the-purpose, smart, and mobile-friendly (Bamwo et al., 2020; Hudson & Barefoot, 2018). Student technology mentorship is the highest trust in students' partnership, where a student becomes the professor's mentor. Academic staff always invest their efforts and dedicated time to develop assessments appropriate to the degree qualification levels. In this venture, the migration onto another or new learning platform does put off some staff, who must redevelop the assessments adapting to the new platform.

In contrast with other types of assessments, quizzes are the most time-consuming and involve the highest efforts, making these "inclusive assessments" with auto-generated marks including comments as feedback for each correct and wrong answer. Imagine the staff's frustration in resurrecting tens of thousands of questions for different modules, when migrating from old StudyNet quizzes to Canvas Quizzes, being not compatible. In higher education, staff's workload is becoming a veridical paradox issue, where staff are expected to carry out additional responsibilities other than delivering regular academic activities (i.e.,

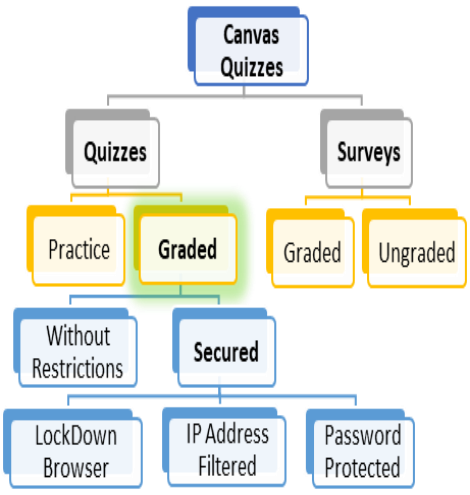
[2] Now known as the School of Physics, Engineering, and Computer Sciences (SPECS)
[3] eduappcenter.com (Edu App Center is provided and maintained by Instructure).

admin, research, and personal development) meeting institutional challenges of various excellence frameworks (Johnson, 2022).

Regardless of the challenges involved in the adaptation of the new LMS platform, pedagogically, the quizzes are most effective only if they are created following cognitive science, which is an interdisciplinary scientific investigation of the mind and intelligence. The quizzes challenge the process of retrieval (i.e., bringing the information to mind), which supports problem-based learning (e.g., learn by doing it), and overall, enhances the learner's ability to apply the information in changing learning environments (e.g., based on physical locations, contexts, and diverse cultures, etc). Thus, the study of quizzes and corresponding generated data can be valuable to identify the issues with individuals and the entire cohort, and a tutor can develop necessary support material, and make necessary changes in future assessments.

Types of Canvas Quizzes, Questions, and their effectiveness

In Canvas, there are two categories of Quizzes, namely: quizzes and surveys, which are further divided into four types as identified in Figure 1(a), and the types of questions and their effective applications are described in Figure 1(b).



(a)

Multiple Choice
True/False
Fill In the Blank
Fill In Multiple Blanks
Multiple Answers
Multiple Dropdowns
Matching
Numerical Answer
Formula Question
Essay Question
File Upload Question
Text (no question)

Type	Effectiveness
T/F	Basic knowledge of a topic, quick proficiency testing.
MCQs, multiple	General recognition of the subject.
Fill in the blanks, multiple, Matching	Checking depth of specific process or rules. Testing chronological, hierarchical, layered fundamentals.
Numerical answer	Testing numeracy skills, accuracy
Formula question	Assessing everyone individually via same question on various STEM's fundamentals, mathematical concepts.
Essay question	Assessing critical thinking, discussion & negotiation skills.
File upload	Similar assessments to Essay type question, but more details of typical method or process that cannot be asked via other Canvas questions.

(b)

Figure 1. (a) Types of Canvas Quizzes, (b) Type of Questions and their effectiveness.

The two types of survey are given as 'Graded' and 'Ungraded' respectively, which can be used for seeking students' experience and satisfaction with specific components (i.e., peer review, delivery of the module, or a course or the entire programme of study). For greater transparency, both types of survey quizzes can be set for anonymous submission using the

Likert scale. These surveys have an additional 'Free Text/Essay' option, which is usually offered to students to suggest any improvement or blog their likes/dislikes other than the points mentioned in the questionnaire. The other category of quizzes is used for formative and summative assessments of learning. The summative quizzes require more scrutiny as these must be offered to students to achieve the desired learning outcomes of a module being linked with the overall programme of studies. After creating a summative assessment (i.e., graded quiz), Canvas automatically sets a column in the Gradebook. If all the questions are set for automatic markings, then the results can be immediately available in the Gradebook. The individual results can be marked and viewed on the 'SpeedGrader' page. Some questions may need manual assessment and marking (i.e., file upload and Essay type questions, respectively).

In graded quizzes, there are protection features to keep students staying on a single browser (e.g., via using Canvas App 'LockDown Browser'), start with a schedule, avoid impersonation or cheating, etc. The tutor can set up an 'access code' that students need to enter before taking the quiz. This means the students will not be able to see or start the quiz until the tutor shares the access code. Furthermore, a tutor can also set a range of IP addresses that are allowed to access the quiz (i.e., within a Lab or on-campus, etc) and students can only take the quiz using computers with an allocated range of IP addresses. For the allocation of IP addresses, the tutor needs some technical assistance from central computing services. The IP address restriction type quizzes are not smartphone friendly as it requires the use of dedicated PCs in the lab.

There are eleven different types of questions as depicted in Figure 1(b), which can be set for both STEM and Applied-STEM disciplines. These can be set to challenge various cognitive levels evaluating the gained skills. In Applied-STEM type quizzes, most objective type questions may deal with the answers, which are based on certain facts and candidates can easily answer by reiterating (i.e., True/False, MCQ, Matching), which are considered to assess the lower order thinking skills (LOTS), based on Blooms' taxonomy, thus do not promote higher order thinking skills (HOTS) (Narayanan & Adithan, 2015; Zaharin et al., 2018; Zaidi et al., 2018). To induce HOTS features, such questions must be asked to conclude by applying the prior knowledge based on logic or computations. Additionally, some questions should be set for assessing students' 'interpretive' or performance measuring skills rather than just for the acquired knowledge.

However, due to time constraints, quiz environmental limitations, and element of chance involved in quizzes, it may not be possible to assess comprehensively a learner using a few types of questions meeting targeted learning outcomes. In this case, the 'Essay' and 'File upload' type questions can be used to effectively test the depth of knowledge, clear understanding of a subject, and critical thinking skills (i.e., market analysis, research on-going industrial and unresolved ethical, environmental and technical business issues, e.g., space junk and Starlink satellites, insurance business of driverless vehicles, humanoid robot

teachers, IoT security and enhancing the quality of services (Jamro, 2021). It is not intended here to cover the implementation of quizzes and questions, however; there are useful video tutorials for academics, tutors, admin, students, and observers in online Canvas resources (Canvas, 2022). In this reference, one can find most of the needed tools and tips from beginner to developer level, developing quiz resources, exploring new quiz features, and more importantly following up on the product roadmap, transitional, and interoperability standards.

Quizzes' Learning Outcomes and Moderation process

The quizzes must be designed with the tutor's planned learning objectives, meeting the desired learning outcomes (LOs) of the module (i.e., MLOs). The LOs are the specific, measurable knowledge, skills, and abilities that the learners are expected to gain by taking the required courses, which are exclusively written for each taught module in the corresponding definite module document (DMD). Several modules LO's are integrated with the LO's of the programme of study (i.e., PLOs), described in the programme specifications. These documents are accredited by PSRBs, centrally reviewed periodically by the university, and updated locally by the Schools. The tutors must make their assessments compliant with these LOs. Therefore, missing out on one small fraction of learning in MLOs can cause a dent in PLO and could create a large vacuum in the learning consequently, devaluing the degree qualification.

In this view, the use of various types of questions and their appropriate use in designing a specific quiz meeting the desired LOs is an artistic creative work of justifying the assessment criteria and meeting the requirements of each LO. The virtual or online quizzes can be different from in-class exams in the operational procedure where besides the invigilators, the academic staff is on standby besides their phones to take students' queries and or verify if there are any mistakes, or errors, or omissions in the quiz paper. In the worst cases of online quizzes, the vulnerable students will get more frustration when they need some clarity about questions, which can be technical, comprehension issues, or could be syntax errors. To avoid such annoyance, a tutor should develop some communication channels via Microsoft Outlook® and Canvas email, Canvas Discussion, personal messages, office phone calls, etc, however; the outlook and Canvas email system require synchronisation and manual enabling of notifications, otherwise, staff may not receive students' queries. This is why the moderation process has an important place in quizzes.

In general, the moderation of assessment is a process that involves instructors, tutors, programme leaders, external examiners, students' mentors, and others involved in a course to discuss student work, award marks on work, give feedback ensuring that marks and feedback are interpreted similarly by all in the moderation group. As a good practice, the moderation must be completed pre- and post-marking, meeting the desired objectives of academic quality assurance. The pre-assessment moderation must analyse the quiz structure, qualification level-appropriate questions' settings, variety of questions,

individualised, customised, and a randomised group of questions, time allowance covering reading and preparation time (RAPT), and more importantly a backup plan if there happens to be an Internet outage or service problems during on-going quiz test. The moderator must also check the availability of all students without any clash in their assignment landscape. The post-marking students' quiz samples must reflect 'inclusive' assessment features to ensure consistency, and equal opportunities for all students (i.e., special needs agreement 'SNA', repeat students, borderline cases, etc.) to demonstrate learners' knowledge, skills, and abilities on the same challenging content (Sadler, 2016; Wilson & Scalise, 2006).

Structure of the Quiz Questions

Pedagogically, the objective-type questions can cover more subject matter than descriptive or essay-type questions. It is important to brief students about various types of quiz questions, their formats, and more importantly tutor's expectations. The questions must be simple, clear and must be consistent with marking, effort, and time constraints. The active students may appreciate the creativity features but the tutor must ensure that the topic is within the context (i.e., not out of course). The quizzes must be designed with the tutor's planned learning objectives, meeting the desired learning outcomes of the module (MLOs). To ensure the quality of a question, the tutor must maintain the main characteristics of an objective-type quiz question, which must contain the necessary elements, including, standard; stimulus; stem; key, and suitable distractors. These elements ensure the quality of the assessment meets the typical learning objectives of a test (Burton et al., 1990; Haladyna et al., 2002; Krish, 2017).

The "standard" establishes the purpose and topic of the assessment item where candidates should be able to read the question and immediately get an illusion of expectation and required action. The standard setting determines a suitable opening or beginning of the problem statement, followed by the problem. The "stimulus" must provide the necessary background information to understand the problem. For STEM disciplines, it can be an expression, equation, or problem whereas for an applied-STEM it can be the background material setting up the challenge. The "stem" identifies the actual problem to solve and must direct correctly to conclude the first two elements (i.e., standards and stimulant). If the stem is irrelevant to the context of the standard, then the assessment will not be suitable for the qualification level. The "key" is the genuine answer that should be unambiguous to skilled learners. Finally, the "distractors" are the other options for the assessment item (i.e., incorrect answers). It is quite a challenge to develop suitable distractors, for example, for students with sufficient KSAs, a distractor should be visible to them, nonetheless; distractors should be challenging enough that they could be seen as the correct answers (i.e., if the learner missed a step or applied a common misconception or tried to guess).

Canvas Quiz Analytics

At any time, the front page in the Canvas module can display a list of things, on the right-hand side, which includes 'New Analytics' as shown in Figure 2. This will show further details of all the Assignments and Quizzes being completed. These analytics contain almost everything a tutor might need for checking the assessment features, students' grades, data plots for writing a module's coursework report for PSRB, module boards, and email contact options to individuals or selected groups for communication and feedback as identified in the upper inset of Figure 2.

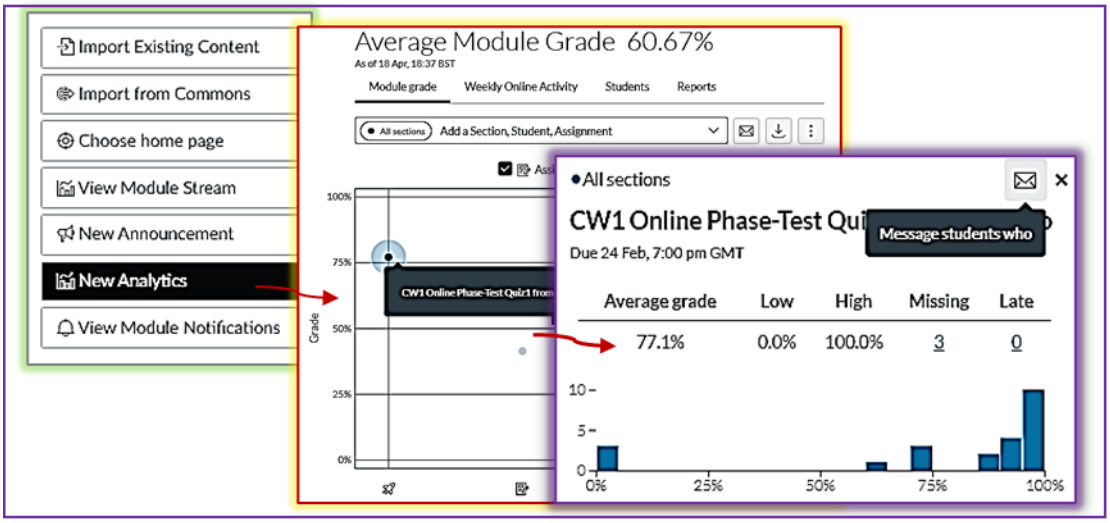


Figure 2. Analytics for Assignments and Quizzes.

Quiz Analysis for Post-marking Moderation

After completing an assessment and releasing marks, the tutor can analyse the quiz statistics based on individuals' performance a Quiz summary is automatically generated showing overall statistical performance with their grade marks as shown in Figure 3. The Y-axis shows the number of students achieving a particular % score displayed as a bar on the X-axis. There is a technical limit for displaying such a plot based on the number of students and or the number of questions.

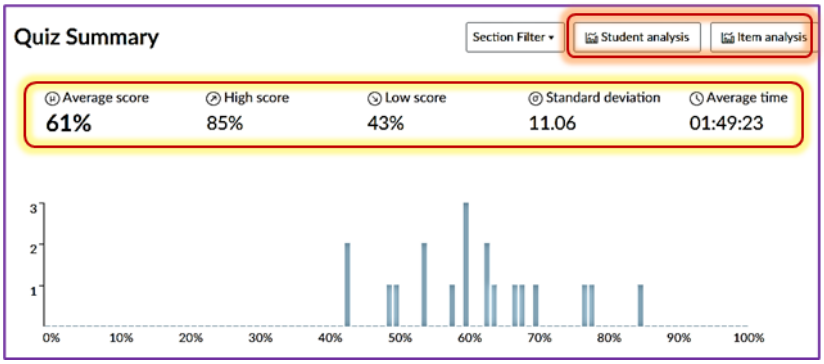


Figure 3. Quiz Summary displaying statistical results.

For example, a quiz with 200 or more questions or 1000 participants will not generate a quiz statistical plot for the visibility feature, however; about 75 or a smaller number of questions will generate a decent quiz plot displaying main statistics. The corresponding data of two useful analyses (i.e., 'Student analysis' and 'Item analysis'), can be downloaded in the CSV (comma separated values) file format of Microsoft Excel® spreadsheets. These provide useful statistical information about quiz analysis, moderation, and class discussion that can be used for reporting to the AQA office about the assessment. Most of the analyses are based on conclusive results comparing various statistical parameters, which are obtained simply based on multiple stages and steps described in the CSV file. For example, the 'Item analysis' contains various useful statistical detail as shown below:

Example of statistical data contained in a typical CSV file

Question ID, Question Title, Answered student count, Variance, Standard Deviation, Alpha score (for the whole exam), Difficulty Index, Discriminatory Index, Top student count (students in the top 27%), Middle student count (students in the middle 46%), Bottom student count (students in the bottom 27%), Quiz question count (total number of quiz questions), Correct student count (number of total students who got the answer right), Wrong student count (number of total students who got the answer wrong), Correct student ratio (ratio of students who got the answer right), Wrong student ratio (ratio of students who got the answer wrong), Correct top student count (students in the top 27% who got the answer right), Correct middle student count (students in the middle 46% who got the answer right), Correct bottom student count (students in the bottom 27% who got the answer right), Point biserial of the correct answer (reliability index), and Point biserial of the first incorrect answer or distractor (followed by the second, etc.).

Each of the aforementioned statistical parameters is based on standard formulas with widely acceptable assumptions. Most of the parameters give various meaningful analyses of individuals, selected groups (top, average, bottom, failed, etc), and also the quality of the quiz and items in the quiz. The fundamental parameter in all the statistical analyses is the variance", which determines the possible number of variations in the data set. It is used to compare, how the selected group's performance varies compared to the sample. Thus, a variance is a simple measure of dispersion (e.g., a process of measuring the distance of each number in the dataset from the mean). The variance is used to calculate the "Standard Deviation", which is a statistical measurement that determines how far a data (or a group) is situated from the mean (i.e., deviates from average). Both these parameters are fundamental to quiz statistics to get analytics of quizzes and questions, therefore; it is essential to understand both.

A simple example is presented here to explain how both these parameters work together.

Example:- Suppose a series of 6 numbers require to determine the standard deviation for a random group. Assume, the numbers to be 4, 18, 34, 2, 26, and 12.

Step #1 Determine the mean (or average) of the numbers:

$$mean = \frac{4 + 18 + 34 + 2 + 26 + 12}{6} = \mathbf{16}$$

Step #2 Subtract the mean from each number, then square the result:

$(4\text{-}16)^2 = 144$

$(18\text{-}16)^2 = 4$

$(34\text{-}16)^2 = 324$

$(2\text{-}16)^2 = 196$

$(26\text{-}16)^2 = 100$

$(12\text{-}16)^2 = 16$

Step #3 Determine the mean of the of the squared values to calculate the variance:

$$variance = \boldsymbol{\mu} = \frac{144+4+324+196+100+16}{6} = \mathbf{130.67}$$

Step #4 Take the square root of the variance to get the Standard Deviation:

$$Standard\ Deviation = \sqrt{136.67} = \mathbf{11.43}$$ It suggests that each number deviates from the mean by 11.43 points on average.

The standard deviation reflects the dispersion of the distribution that is described by a bell-shaped 'Gaussian' profile as given in Figure 4. A low standard deviation indicates that the data value is located close to the mean (i.e., expected value), whereas a high standard deviation indicates that the data values are spread out over a wide range. Three different standard deviations with a normal distribution profile are shown in Figure 4.

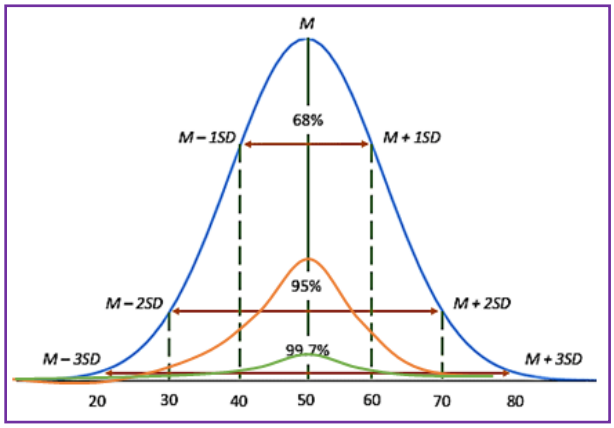


Figure 4. Standard Deviation from a mean value.

The blue bell-shaped curve demonstrates the one standard deviation from the mean value of 50 with standard deviation, it will produce empirical values between 40 and 60, that is about 68%. The 'two standard deviations' will press the height down (i.e., reducing

amplitude) and consequently disperse (i.e., flattening on the x-axis). Now the data has mean values located between 30 and 70 (i.e., about 95%). The 'three standard deviations' will result in 99.7% further deviating the expected data values from the mean within a diverse range between 20 and 80. Fundamentally, the curve with the lowest standard deviation has a high peak and a small spread, while the curve with the highest standard deviation is flatter and data is widespread (i.e., it contains the least expected values).

There are also useful statistics that determine the quality of questions that can be used to evaluate hidden features of questions. For example, for some tutors, a multiple-choice question (MCQ) may appear as a simple selection of options but from a psychometric point, it is a complex process, which demands to develop of reasonable but incorrect options (i.e., distractors) for each item is a subject of research for content specialists (Shin et al., 2019). The distractors undergo a test based on students' most common misconceptions, to qualitatively identify whether all the alternative options functioned as intended. The poor distractors will appear as irrelevant options in the True/False or MCQs or MRQs and must be removed from the options of the quiz test. The quality of distractors is measured by using the point-biserial correlation coefficient that relates observed item responses and is particularly used when one set of data is dichotomous (i.e., separated into two mutually exclusive or contradictory optional groups, e.g., right and wrong). For a quiz question, it calculates the values based on correct and incorrect responses. Additionally, Canvas quizzes also deal with the distractors to identify the quality or efficiency of each distractor. Based on Pearson's correlation coefficient the point biserial correlation coefficient can be calculated after finding the relationship between correct and incorrect answers given by the participants (Papenberg & Musch, 2017):

$$r_{pb} = \frac{\overline{Y_1} - \overline{Y_0}}{S_y} \sqrt{\frac{n_0\, n_1}{n^2}} \qquad \ldots\ldots\ldots\ldots \qquad \text{Eq. 2}$$

The term 'y1' is the mean test score of students correctly identifying the solution (i.e., group 1), whereas 'y0' is the mean score of students choosing the distractor (i.e., group 0). The parameter 'Sy' is the standard deviation of all test scores. The 'n1' is the number of students choosing the solution (i.e., group 1), and 'n0' is the number of test takers choosing the distractor (i.e., group 0), and the 'n' is the total number of students (i.e., both groups 0 and 1).

The other two useful parameters are the difficulty and discrimination indices, which are also determined from the data produced as an excel spreadsheet. The 'Difficulty Index' parameter 'r' (also known as the p-value) determines the challenge (i.e., if the question was too easy or too difficult). It can be expressed based on the data obtained for the top and bottom groups, given by an expression (Gronlund & Linn, 1990):

$$\rho = \frac{Tc + Bc}{n} \qquad \ldots\ldots\ldots\ldots \qquad \text{Eq. 3}$$

where the terms, 'TC' and 'BC' are the participants, who answered correctly from the Top third high achieving group, and Bottom third low achieving groups, respectively. The 'n' represents the number of total participants in both groups. The p-value should lie anywhere within a range of ±1. A value of below 0.25 suggests that the question is very difficult, and the tutor must review the question for complications, missing details, and confusing language, and must highlight that topic for more students' support, or should check if the question is level appropriate or not. On the other hand, the values above 0.75 are considered very easy and must not be set for high marks or can be removed from the future summative tests. The values between 0.25 and 0.75 are considered average or moderate.

The Discrimination Index determines the quality of the cohort describing how well it discriminated between high and low achievers. It inspects each quiz item and evaluates how well an item can differentiate between good and less able students. Overall, it provides a performance comparison of stronger vs weaker students in the quiz. Mathematically, it can be written as an equation:

$$DI = \frac{Tc - Bc}{\frac{1}{2}(n)} \quad \text{.....} \quad \text{Eq. 4}$$

Theoretically, a DI could have any value within ±1, where the closest to +1 is more effective. Nonetheless, it is not an absolute measure but does help mostly identify the effectiveness of the most objective type of quiz questions where students are offered options with distractors (e.g., T/F, MCQs, MRQS, and Matching). In Canvas quizzes, +0.24 or lower values are considered as the poor scores, whereas a higher than +0.25 is a better performance indicator. Interestingly, a DI of zero shows all students getting the quiz question right (or wrong). Reusing the high DI values in the next quizzes is recommended, as these possess higher cognitive thinking abilities (i.e., HOTS). The poor discriminative questions must be either removed or revised for improvement. The DI can be found in the 'Question Breakdown' of a quiz providing a comprehensive report of the respondents with corresponding percentages of correct, and incorrect answers, thereby displaying a 'Discrimination Index' as shown in Figure 5.

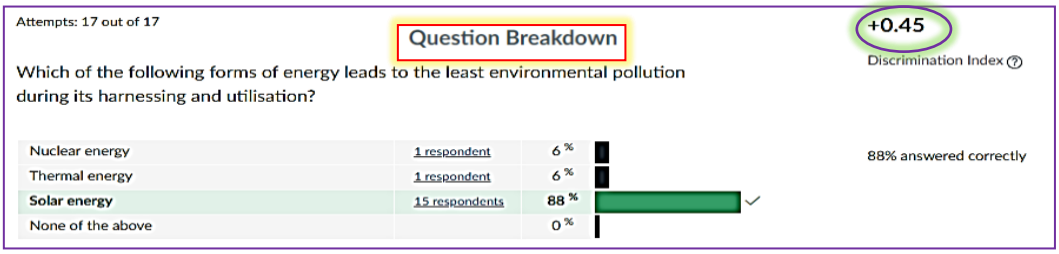


Figure 5. Quiz question showing Discrimination Index value.

There are various possibilities to analyse the data confirming the statistical assumptions of the surveys and quizzes. For example, to determine if a difference between "observed" and "expected" data is purely due to a chance, or if there exists a relationship between the

variables used in the test. In quizzes, this can be used to study the quality factors with objective-type questions (e.g., T/F, MCQs, MRQs, Matching, etc). One such analysis is always performed via a test is known as Chi-square ($\chi 2$) test, which can be given as a simple expression (Comrey & Lee, 2006):

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \ldots\ldots\ldots\ldots \qquad \text{Eq. 5}$$

where the terms '*O*' and '*E*' represent the observed and expected values, respectively. The range is determined from *i* =1 to '*n*'. There can be three main types of Chi-square tests, the test of goodness of fit (e.g., if data fits a particular distribution), the test of independence (e.g., if the test is independent of two factors), and the test for homogeneity (e.g., if the differences are consistent).

Overall, the statistical analysis after completing an assessment offers a useful SWOT (Strengths, Weaknesses, Opportunities, Threats) that determines the behaviour, confidence, and sustained learning of the participants, respectively. These are essentials of a "pedagogical framework" that determines the integrated LT&A values and instructor's consistent practice and support within the institution (AlMarwani, 2020; Granić et al., 2009; Longhurst et al., 2020). In this view, the quizzes analyses can effectively be used to diagnose students' behaviour, incentives and pressures to engage in cheating misconduct, or opportunities that arise from different patterns of online learning and assessment (i.e., pre, post-COVID-19, and during the pandemic) (White, 2020). In response to various SWOT analyses on online quizzes and assessments, a common academic debate suggests that the educational institutes want students to avoid cheating practices in online quizzes, thus they must educate academic members, set resilient protections, and encourage good practice in LT&A and feedback. It is therefore recommended that the summative quizzes with more marks allowance must be seriously commissioned by the relevant LT&A team members. The central and schools' AQA, L&T, and HR teams must involve in the staff's development opportunities (e.g., peer teaching observation, internal moderation, and continuous professional academic development; CPAD) encouraging young academics to make efficient use of quiz statistics for their teaching modules.

Conclusions and Recommendations

Good-practised features of Canvas Quizzes were explored concerning assessment and feedback, covering the statistical parameters that a tutor might require to design quizzes for students or moderate the quizzes for other staff. Useful analytical features of the Canvas quizzes were scrutinised that a tutor, moderator, or programme leader can use to report to the office of the AQA. It is recommended that the AQA team encourage programme leaders and academic staff to produce psychometric details of their students' assessments, in particular quizzes, and share their good-practised examples in module boards, learning and teaching workshops, seminars and L&T conferences. These must be regarded as similar to "scholarly activities" or research contributions (i.e., REF for subjective research). If staff is

deprived of time, resources, and efforts in developing alternative assessments, making the learning interesting and fun, then staff will not invest their full potential in learning new LT&A and feedback tools. The learning and teaching Innovation centre (LTIC), learning technologists and academics must collaborate on the staff's development opportunities, e.g., continuous professional academic development (CPAD) training academics of STEM and applied-STEM on quizzes statistics for effective use in scholarly activities.

The University of Hertfordshire has completed the initial transaction of the migration to Canvas and there are still more ongoing digitalisation, artificial intelligence and deep learning challenges ahead. The Canvas quizzes will be revised in June 2024 and additionally more 3rd party quiz apps will be added to Canvas, which can effectively produce the required data for writing reports meeting the accreditation bodies' strict requirements. Additionally, the academics must be encouraged to exchange their queries and concerns with the Canvas champions and learning technologies team, who support the implementation of the required technology tools, and apps and continue enhancing the learning and assessments technologies.

References

A Kaklij, V., Shah, V., Kunal, M., & Mandawkar, M. (2020). Microlearning based content-curation using Artificial Intelligence for Learning Experience Platform: A Survey. *Shah, Mr. Kunal and Mandawkar, Mr. Umakant, Microlearning Based Content-Curation Using Artificial Intelligence for Learning Experience Platform: A Survey (August 19, 2020). IJRAR-International Journal of Research and Analytical Reviews (IJRAR), E-ISSN*, 1269–2348.

AlMarwani, M. (2020). Pedagogical potential of SWOT analysis: An approach to teaching critical thinking. *Thinking Skills and Creativity*, *38*, 100741.

Atif, A., Jha, M., Richards, D., & Bilgin, A. A. (2021). Artificial Intelligence (AI)-enabled remote learning and teaching using Pedagogical Conversational Agents and Learning Analytics. In *Intelligent Systems and Learning Data Analytics in Online Education* (pp. 3–29). Elsevier.

Bali, M. (2014). MOOC pedagogy: Gleaning good practice from existing MOOCs. *Journal of Online Learning and Teaching*, *10*(1), 44.

Bamwo, L., Clarkson, S., & Benford-Brown, T. (2020). Implementing the Guided Learner Journey: A case study of the evolution of institutional wide partnerships, through a set of pedagogic principles. *The Journal of Educational Innovation, Partnership and Change*, *6*(1).

Bull, J., & McKenna, C. (2003). *A blueprint for computer-assisted assessment*. Routledge.

Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1990). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Brigham Young University. Department of Instructional Science.

Canvas. (2022). *Canvas Community Video Guide*. https://community.canvaslms.com/t5/Video-Guide/tkb-p/videos

Canvas_Community. (2021). *Canvas New Quizzes Feature Comparison*. https://docs.google.com/document/d/11nSS2EP0UpSM6dcuEFnoF-hC6lyqWbE9JSHELNmfG2A/edit

Comrey, A. L., & Lee, H. B. (2006). *Chapter 11: Analysing Frequency Data with Chi Square, in Elementary Statistics: A Problem Solving Approach* (4th ed.). Lulu. com.

DeCoito, I., & Estaiteyeh, M. (2022). Online teaching during the COVID-19 pandemic: exploring science/STEM teachers' curriculum and assessment practices in Canada. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 1–18.

Diaz, J. B. B., & Young, S. F. (2021). The future is here: A benchmark study of digitally enabled assessment and development tools. *Consulting Psychology Journal: Practice and Research*.

Granić, A., Mifsud, C., & Ćukušić, M. (2009). Design, implementation and validation of a Europe-wide pedagogical framework for e-learning. *Computers & Education*, *53*(4), 1052–1081.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and Evaluation in Teaching.* Prentice Hall.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–333.

Hudson, P., & Barefoot, H. (2018). *Developing the Guided Learner Journey*.

IBL_News. (2019, August 30). *LMS Market Won't Grow, Continuing To Be Dominated By The Big Four*. https://iblnews.org/the-lms-market-doesnt-grow-and-continues-to-be-dominated-by-the-big-four/

IMS-Global. (2019). *Learning Tools Interoperability LEARNING TOOLS: 1.3 AND LTI ADVANTAGE ARE TRANSFORMING DIGITAL LEARNING*. IMS Global. https://www.imsglobal.org/spec/lti/v1p3/impl

Jamro, M. Y. (2017). Engaging students of Engineering & Technology in positive Group Dynamics using randomly generated Instant Virtual Group (IVG) quizzes. *International Journal for Innovation Education and Research*, *5*(10), 1–35. https://doi.org/10.31686

Jamro, M.Y. (2021). IoT Security with QoS: Game changer for Industry and STEM Education. 2021 International Carnahan Conference on Security Technology (ICCST), 1–4.

Jefferies, A., Cubric, M., & Russell, M. (2013). Enhancing learning and teaching using electronic voting systems–The development of a framework for an institutional approach for their introduction. In *Increasing Student Engagement and Retention Using Classroom Technologies: Classroom Response Systems and Mediated Discourse Technologies*. Emerald Group Publishing Limited.

Johnson, M. T. (2022). The knowledge exchange framework: understanding parameters and the capacity for transformative engagement. *Studies in Higher Education*, *47*(1), 194–211.

Justin. (2021). *The UK LMS Market is Changing*. Listedtech. https://listedtech.com/blog/the-uk-lms-market-is-changing/

Justin. (2022). *Instructure's Impressive Implementations Record*. Listedtech. https://listedtech.com/blog/instructures-impressive-implementations-record/

Kiran. (2020). *Moodle vs Canvas 2021- Which LMS is better?* (Edwiser). https://edwiser.org/blog/moodle-vs-canvas/

Krish, N. (2017). *The Anatomy Of Assessment: 5 Elements Of A Quality Multiple Choice Question, Assessment Resources For Teachers*. Teachthought.Com. https://www.teachthought.com/pedagogy/quality-multiple-choice/

Longhurst, G. J., Stone, D. M., Dulohery, K., Scully, D., Campbell, T., & Smith, C. F. (2020). Strength, weakness, opportunity, threat (SWOT) analysis of the adaptations to anatomical education in the United Kingdom and Republic of Ireland in response to the Covid-19 pandemic. *Anatomical Sciences Education*, *13*(3), 301–311.

Narayanan, S., & Adithan, M. (2015). Analysis of question papers in engineering courses with respect to HOTS (Higher Order Thinking Skills). *American Journal of Engineering Education (AJEE)*, *6*(1), 1–10.

Öztürk, Y. E., & Gürler, İ. (2020). Evaluation of Moodle, Canvas, Blackboard, and Open EdX. In *ICT-based assessment, methods, and programs in tertiary education* (pp. 363–382). IGI Global.

Papenberg, M., & Musch, J. (2017). Of small beauties and large beasts: The quality of distractors on multiple-choice tests is more important than their quantity. *Applied Measurement in Education*, *30*(4), 273–286.

Paynter, K., & Barnes, J. (2021). Moving from Blackboard to Canvas: What the research says, plus two professors' experiences. *International Journal on E-Learning*, *20*(1), 5–16.

Sadler, D. R. (2016). Three in-course assessment reforms to improve higher education learning outcomes. *Assessment & Evaluation in Higher Education*, *41*(7), 1081–1099.

Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. John Wiley & Sons.

Shah, D. (2019). *Year of MOOC-based Degrees: A Review of MOOC Stats and Trends in 2018,* *https://www.classcentral.com/report/moocs-stats-and-trends-2018/*.

Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distractor development using topic modeling approaches. *Frontiers in Psychology*, *10*, 825.

Sim, S. P.-L., Sim, H. P.-K., & Quah, C.-S. (2021). Online learning: A post COVID-19 alternative pedagogy for university students. *Asian Journal of University Education*, *16*(4), 137–151.

Swerzenski, J. D. (2021). Critically analyzing the online classroom: Blackboard, Moodle, Canvas, and the pedagogy they produce. *Journal of Communication Pedagogy*, *4*, 51–69.

Telles-Langdon, D. M. (2020). Transitioning University Courses Online in Response to COVID-19. *Journal of Teaching and Learning*, *14*(1). https://doi.org/10.22329/jtl.v14i1.6262

Tuah, N. A. A., & Naing, L. (2021). Is online assessment in higher education institutions during COVID-19 pandemic reliable? *Siriraj Medical Journal*, *73*(1), 61–68.

van Miegroet, H., Glass, C., Callister, R. R., & Sullivan, K. (2019). Unclogging the pipeline: Advancement to full professor in academic STEM. *Equality, Diversity and Inclusion: An International Journal*.

White, A. (2020). May you live in interesting times: a reflection on academic integrity and accounting assessment during COVID19 and online learning. *Accounting Research Journal*.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher Education*, *52*(4), 635–663.

Yıldırım, B. (2020). MOOCs in STEM Education: Teacher Preparation and Views. *Technology, Knowledge and Learning*, 1–26.

Zaharin, N. L., Sharif, S., & Mariappan, M. (2018). Computational thinking: A strategy for developing problem solving skills and Higher Order Thinking Skills (HOTs). *Int. J. Acad. Res. Bus. Soc. Sci*, *8*(10).

Zaidi, N. L. B., Grob, K. L., Monrad, S. M., Kurtz, J. B., Tai, A., Ahmed, A. Z., Gruppen, L. D., & Santen, S. A. (2018). Pushing critical thinking skills with multiple-choice questions: does Bloom's taxonomy work? *Academic Medicine*, *93*(6), 856–859.