The Canadian Journal for the Scholarship of Teaching and Learning

Volume 15 | Issue 2 Article 10

Summer 08-31-2024

# Student-generated Multiple-Choice Questions: A Java and Web-Based Tool for Students to Create Multiple Choice Tests

Larry Katz
*University of Calgary*, katz@ucalgary.ca
Dave Carlgren
*SAIT,* dave.carlgren@sait.ca
Cory Wright-Maley
*St. Mary's University*, Cory.WrightMaley@stmu.ca
Megan Hallam
*Ambrose University*, mhallam@ambrose.edu
Joan Forder
*University of Saskatchewan*, joan.forder@usask.ca
Danielle Milner
daniellemilner9@gmail.com
Lisa Finestone
*University of Calgary*, lisa.finestone@ucalgary.ca


https://doi.org/10.5206/cjsotlrcacea.2024.2.16625

Recommended Citation
Katz, L., Carlgren, D., Wright-Maley, C., Hallam, M., Forder, J., Milner, D., & Finestone, L. (2024). Student-generated multiple-choice questions: A Java and web-based tool for students to create multiple choice tests. *The Canadian Journal for the Scholarship of Teaching and Learning, 15*(2). https://doi.org/10.5206/cjsotlrcacea.2024.2.16625

# Student-generated Multiple-Choice Questions: A Java and Web-Based Tool for Students to Create Multiple Choice Tests

**Abstract**
Student-generated questions can be an effective study technique to improve active learning, metacognitive skills, and performance on examinations. Students have shown greater success when assessed using peer-made study questions than when studying without questions. In three semesters of a kinesiology research methods course students were taught how to write high-quality multiple-choice questions that addressed course objectives and Bloom's Taxonomy. Students were given a graded assignment to write three multiple-choice questions using Makequiz, a Java and web-based tool for helping students generate multiple choice questions. Student- generated questions that were rated as good quality ($n$ = 169-245) were provided to the students as a study resource prior to the final exam. Of those study questions, 40 were selected each semester to be on the final exam. Students performed significantly better on student-written questions than instructor-written questions on the final exam in Class A ($p < .05$) and in Class C ($p < .05$). A majority of students felt this assignment was a worthwhile component of the course, voting to keep Makequiz in the curriculum (Class A: 52.6%, Class B: 62.3%, Class C: 58.3%) or to modify Makequiz (Class A: 25.3%, Class B: 14.5%, Class C: 18.1%). Many of the students stated it was the most valuable assignment of the course (32% Class B and Class C). Makequiz is, therefore, a recommended study tool for students. Comparisons are also made with PeerWise, an online platform for creating and sharing MCQ and feedback. Further investigation is required to measure the impact Makequiz has on learning, metacognitive skills, and anxiety levels before test-taking.

Les questions générées par les étudiants et les étudiantes peuvent être une technique d'apprentissage efficace pour améliorer l'apprentissage actif, les compétences métacognitives et les performances aux examens. Les étudiants et les étudiantes ont obtenu de meilleurs résultats lorsqu'ils ont été évalués à l'aide de questions créées par leurs pairs que lorsqu'ils ont étudié sans questions. Au cours de trois semestres d'un cours de méthodes de recherche en kinésiologie, les étudiants et les étudiantes ont appris à rédiger des questions à choix multiples de haute qualité qui répondaient aux objectifs du cours et à la taxonomie de Bloom. Les étudiants et les étudiantes ont été notés pour rédiger trois questions à choix multiples à l'aide de Makequiz, un outil Java basé sur le web pour aider les étudiants et les étudiantes à générer des questions à choix multiples. Les questions générées par les étudiants et les étudiantes et jugées de bonne qualité ($n$ = 169-245) ont été fournies aux étudiants et aux étudiantes comme ressource d'apprentissage avant l'examen final. Parmi ces questions, 40 ont été sélectionnées chaque semestre pour figurer dans l'examen final. Les étudiants et les étudiantes ont obtenu de bien meilleurs résultats aux questions rédigées par les étudiants et les étudiantes qu'aux questions rédigées par l'enseignant ou l'enseignante lors de l'examen final dans la classe A ($p < 0,05$) et dans la classe C ($p < 0,05$). Une majorité d'étudiants et d'étudiantes ont estimé que ce travail était une composante intéressante du cours et ont voté pour le maintien de Makequiz dans le programme (classe A : 52,6%, classe B : 62,3%, classe C : 58,3%) ou pour la modification de Makequiz (classe A : 25,3%, classe B : 14,5%, classe C : 18,1%). De nombreux étudiants et de nombreuses étudiantes ont déclaré qu'il s'agissait de l'exercice le plus utile du cours (32% en classe B et classe C). Makequiz est donc un outil d'apprentissage recommandé pour les étudiants et les étudiantes. Des comparaisons sont également faites avec PeerWise, une plate-forme en ligne pour la création et le partage de questions à choix multiple et de rétroaction. Des recherches supplémentaires sont nécessaires pour mesurer l'impact de Makequiz sur l'apprentissage, les compétences métacognitives et les niveaux d'anxiété avant les examens.

**Keywords**
multiple choice questions, computer assisted learning, student-generated questions, Bloom's taxonomy, metacognition; questions à choix multiples, apprentissage assisté par ordinateur, questions générées par les étudiants et les étudiantes, taxonomie de Bloom, métacognition

This research paper/Rapport de recherche is available in The Canadian Journal for the Scholarship of Teaching and Learning: https://doi.org/10.5206/cjsotlrcacea.2024.2.16625

With the trend toward larger class sizes and a subsequent need for efficient marking methods, multiple choice questions (MCQs) have become widely used in university testing, often accounting for a substantial portion of the course grade (Mavis et al., 2001; Parkes & Zimmaro, 2016). A typical MCQ consists of a question stem and four or five possible answers. The correct option is referred to as the keyed option and the others are distractors.

Despite their ubiquity, MCQs are often criticized for being of poor quality (DiBattista & Kurzawa, 2011; Stanley, 2021; Zimmaro, 2004). Zimmaro (2004) reported that many MCQs contain poor stems and/or non-effective distractors. One study found that 45% of the distractors were flawed as they were chosen by fewer than 5% of students (DiBattista & Kurzawa, 2011). Similarly dismal results have been found by others examining higher education multiple choice exams in contexts such as medical school (Ware & Vik, 2009) and nursing (Tarrant et al., 2009). Stanley (2021) boldly states, "[h]ow many times in the real world are we called upon in to take a multiple-choice test?" (p.1).

However, Bacon (2003) reported that well-constructed multiple-choice examinations can yield test scores that are at least as reliable as those produced by a constructed-response test. When dysfunctional distractors were either replaced or deleted, the reliability of the questions increases (Cizek & O'Day, 1994). Thus, careful consideration in the generation of MCQs is critically important.

## Bloom's Taxonomy

Bloom's Taxonomy is a hierarchical classification of learning objectives that education experts utilize to guide and measure depth of learning (Bloom, 1956; Morton & Colbert-Getz, 2017; Thompson & O'Loughlin, 2015). Using Bloom's Taxonomy, Lord & Baviskar (2007) found that courses which focused on teaching and evaluating detailed, factual content failed to provide students with opportunities to develop higher level thinking. It is said that MCQs most frequently attend to lower levels of thinking, such as the Knowledge or Comprehension levels of Bloom's Taxonomy (Denny et al., 2008a; Walsh & Seldomridge, 2006).

Students themselves perceive that a lower level of intellectual processing is assessed by MCQs, leading to a shallower learning approach when compared to preparing an assignment essay (Scouller, 1998).

More research has since been done demonstrating that MCQs can be thoughtfully designed to incorporate all domains of Bloom's Taxonomy to accurately assess critical thinking (Bates et al., 2014; Gonzalez-Cabezas et al., 2015; Kim et al., 2012; Zaidi et al., 2018). This is further supported with Zaidi and colleagues' (2017) finding that higher-performing students do significantly better on higher-level Bloom's Taxonomy MCQs than do lower-performing students.

## Student-Generated Questions

Analysis of student-generated versus faculty-generated MCQs showed no comparable difference in quality (McLeod & Snell, 1996). However, professionals think differently than the students for whom they write questions. They tend to automatically ignore irrelevant information and associate necessary information, allocating attention in ways students may not (Ding et al., 2009). As such, avoiding misinterpretation and encouraging active thinking may be better achieved by having students generate their own questions. In addition, Zaidi et al. (2018) reported "that faculty content experts (generally the question writer) and student novice learners (the examinee)

approach MCQs in different ways—largely because of differences in foundational knowledge, which influence the cognitive levels being used to address an MCQ" (p. 857).

Several scholars have studied interventions designed to help students and faculty develop higher quality MCQs that aim to test higher-order thinking skills (e.g., Berry & Chew, 2008; Foote, 1998; King, 1992; Lord & Baviskar, 2007; Rosenshine et al., 1996; Zaidi et al., 2018). Students' questions give insight into what they understand, what they think is important, facilitate deeper thought into material, and guide teachers to student needs (Foster, 2011).

Additionally, student engagement in course material is facilitated through student creation of questions due to enjoyment of such activity (Gooi & Sommerfeld, 2015). Student-generated questions can be an effective study technique (Craft, 2017; Foos et al., 1994; Foos, 1989; Green 1997; Kay, et al., 2020) and strategy for promoting deeper learning (Chickering & Gamson, 1987; Fellenz, 2004). This was found to be true also by providing marks to students regarding their own MCQs and having some of these questions incorporated into summative assessments (Chickering & Gamson, 1987; Fellenz, 2004). It has been pointed out that deeper learning requires focus on learning relationships between items rather than on recalling disconnected true-false items (Draper, 2009). Thus, it is necessary to provide students not only with practice in developing MCQs as a strategy for improving learning, but also to ensure they have opportunities to see and build connections across course content. Beyond generating their own questions, sharing these questions to allow students to study with peer-generated questions has been shown to be beneficial for learning (Foos et al., 1994; Gooi & Sommerfeld, 2015; Green, 1997). When studying with peer-generated questions, students scored significantly higher on summative assessments than those who studied without the use of these questions (Foos et al., 1994). Furthermore, student dedication to a course increased although test anxiety was simultaneously reduced (Green, 1997).

King (1992) conducted a number of studies to evaluate the effectiveness of a guided student-generated questioning strategy, designed to facilitate students' higher-order thinking. Although the internal validity of his studies has been questioned by others (e.g., Foote 1998), King ascertained that such strategies promote critical thinking and a higher degree of learner autonomy. These findings are corroborated by others who have noted that students' generation of questions results in improved confidence and active learning (Craft et al., 2017), comprehension and deeper learning of material as well as higher-order thinking (Draper, 2009; Lord & Baviskar, 2007; Rosenshine et al., 1996). When involving students in question creation, they should be encouraged to formulate questions of mid- and higher-thinking levels of this taxonomy (Lord & Baviskar, 2007). Incorporating this study strategy into classes increases collaborative interactions and student engagement in course material (Song et al., 2017).

## Metacognition

Metcalfe (2009) described that metacognitive skills involve judging one's proximal zone of knowledge and expanding that zone. Studying can therefore be made more effective with improved metacognition (Metcalfe, 2009). The ability to improve between pre- and post-tests has been found to be dependent on metacognitive knowledge for adult students with high levels of monitoring ability (Schraw, 1994), as has student propensity to integrate study strategies (de Carvalho Filho, 2010).

However, one cannot assume that learners will develop their capacity to use metacognitive tools on their own. Pressley and Ghatala (1990) found a disjunction between adult learners' perceptions of performance and actual performance on comprehension test items. In their study,

learners overestimated their accuracy, thus limiting their ability to recognize errors or reconsider questions. Specific training is needed to properly use metacognitive strategies (Huff & Nietfeld, 2009).

Justifying an answer during studying yields better results in testing (Koretsky et al., 2016) and feedback after answering a question has been shown to improve metacognitive judgements as well as test performance (Callender et al., 2016; Egelandsdal & Krumsvik, 2017). Generation of questions, studying with peer-generated questions, and receiving feedback during such studying are all viable ways to improve metacognition, thereby facilitating learning.

## Makequiz Program

Computer programs are available for students to generate their own questions and when students evaluate questions created by peers, they engage in higher-order thinking (Denny et al., 2008c). Denny and associates reported that student participation in question generation and evaluation exercises lead to deeper learning for the participating students, across all of the major topics in the curriculum (Denny et al., 2008c; Denny et al., 2009).

Makequiz (Calgary, Canada) is a Java and web-based computer program developed by the first author as a platform for students to design their own MCQs (Figure 1) and study peer-written MCQs (Figure 2). Question-authors write feedback for each multiple-choice answer explaining why each distractor is incorrect and why each keyed answer is correct. This feedback is immediately presented to students upon selecting and confirming an answer. The purpose of Makequiz is to 1) increase student interaction in higher education classes, 2) deepen learning by writing questions, distractors, and effective explanations, and 3) improve metacognitive skills by receiving immediate feedback. Generating thoughtful questions and studying peer-written Questions may help learners actively approach their studies, allowing them to both undertake the endeavour, and reflect upon it effectively (Flavell & Wellman, 1977).

The present study explores the value of using Makequiz in university classes. In preparation for a final exam, students used Makequiz to create their own MCQs and then received questions written by peers with which to study. Student performance on peer-written questions relative to instructor-written questions was measured and student perspectives on Makequiz were evaluated.

**Figure 1**

*Creation of an MCQ Using Makequiz with Correct Answer Keyed, Distractors, and Explanations for Each*

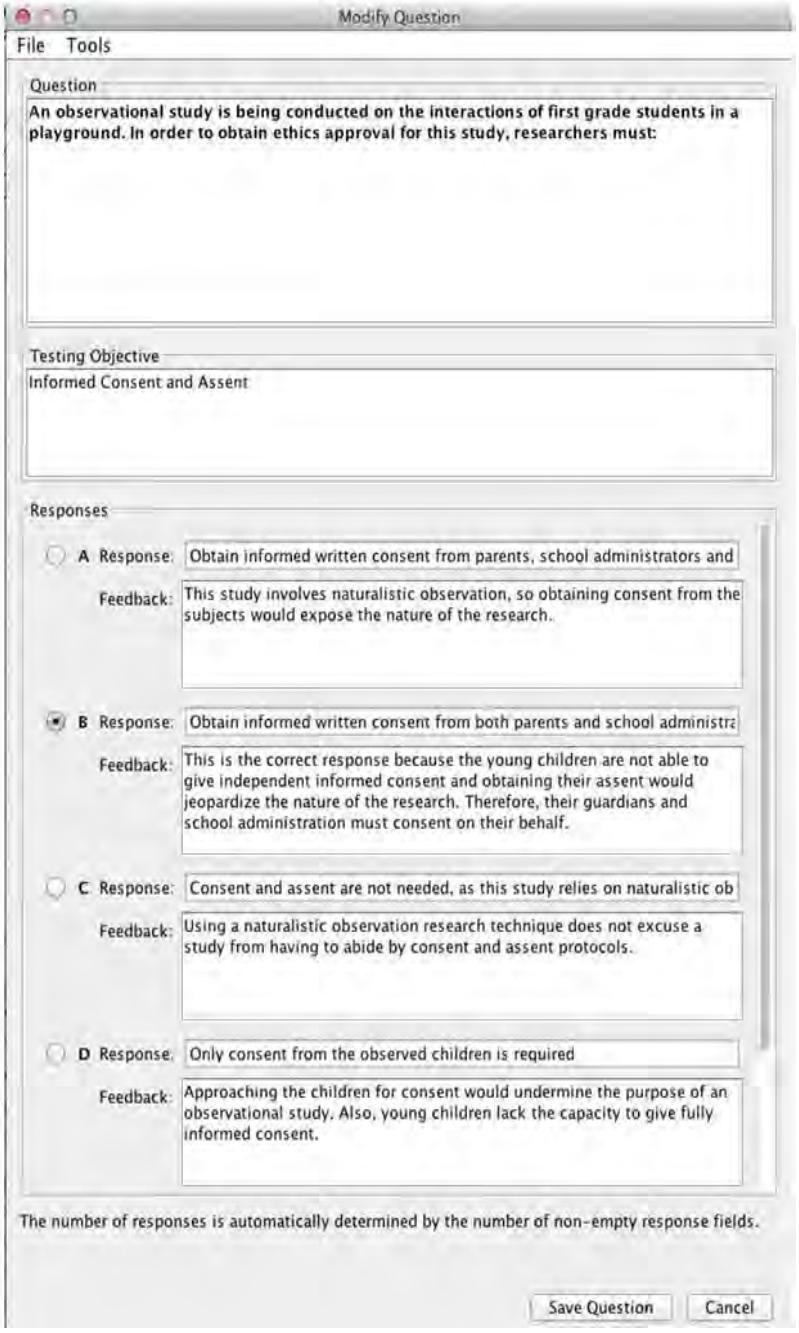
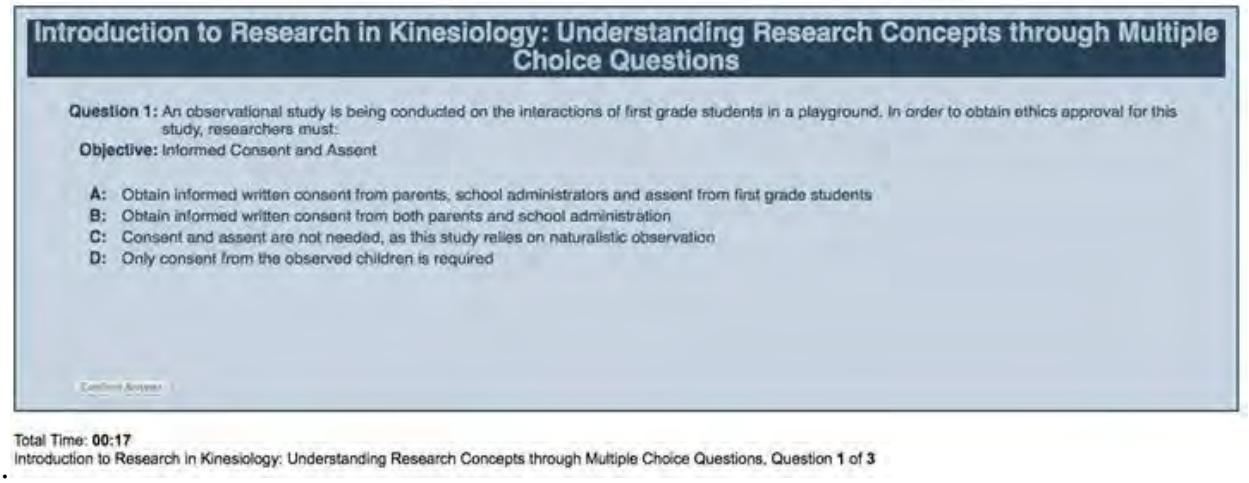


**Figure 2**

*Studying with Peer-written MCQs in Makequiz.*




## Method

### Participants

Participants included three multi-level undergraduate kinesiology student cohorts enrolled in Introduction to Research Methods over two years; 118 students from Class A, 111 students from Class B, and 111 students from Class C. As an introductory level course, the largest number of participants came from the freshman class (Class A – 52%, Class B – 47%, Class C – 50%); however, all undergraduate levels were represented in each of the three classes. In two classes, the majority of students came from the Faculty of Kinesiology (Class A – 77%, Class B – 48%, Class C – 71%). The remaining students in these classes came from the faculties of Science, Art, and Business. After consultation with the Conjoint Faculty Research Board Ethics at the university, formal ethics approval was not required because the student information was anonymous, only group data was used, and student feedback was anonymous. Students were also asked to complete an optional and anonymous course evaluation form. A copy of this assessment is included in Appendix A. Additionally, this paper deals primarily with the processes of learning with student-generated multiple-choice questions and not the specifics of the students themselves.

### Procedure

#### *Question-Writing Training*

A specialist from the university's Teaching and Learning Centre was enlisted to teach a lesson on writing MCQs. This lesson addressed how to avoid common flaws in the design of MCQs that cue readers to the answers, the qualities of good and bad MCQs, and how to write higher-order thinking MCQs within the framework of Bloom's Taxonomy (See Appendix A). Students were provided with a list of Do's and Don'ts to guide their question-writing (Table 1).

**Table 1**

*Items Evaluated for Do's, Don'ts, and Overall Score*

| Do's | Don'ts | Overall Score |
|---|---|---|
| • All information in the stem is important<br>• Stem is clear & specific<br>• Question is unambiguous<br>• Answer is not dependent<br>• Significant content or skill required<br>• Question is of Higher or Middle Order of Bloom's Taxonomy | • Clang Associations<br>• Exceptionally long answer<br>• Grammar inconsistencies<br>• Absurd/implausible answer<br>• Absolute answers<br>• All of the above/none of the above<br>• Negatively phrased stem<br>• Overlap in items<br>• True distractor does not answer<br>• Deliberately tricky question | • Bloom's level identified<br>• Middle-Higher Order<br>• 4 alternative answers present<br>• Question objective stated<br>• Quality of stem<br>• Quality of distractors<br>• Quality of distractor feedback<br>• Correct answer feedback |

In the following class, students were taught how to use the Makequiz program to construct MCQs. They were provided with a free copy of Makequiz to create MCQs and write feedback on correct and incorrect answers. Although these two lectures were given in all three semesters, more in-depth instruction was provided during the last two semesters (Class B, Class C) than in the first semester (Class A) due to a change in the availability of specialists in MCQ instruction. Classes B and C received instruction that included a more detailed presentation than the first semester and more examples of variable question quality as well as interactive critical analysis of question quality.

During the specialist's presentation, all students were able to test and get feedback on their knowledge regarding good MCQs using SMART Technologies' Student Response System (Calgary, Canada).

### *Assignment*

Students were assigned to create three MCQs in Makequiz, based on one of several course objectives. These questions were required to employ Bloom's Taxonomy (1956). From a list provided at the beginning of the semester, students selected three distinct class objectives upon which to base their questions, and each objective could be selected by up to three students.

There was a list of 120 objectives in Class A, but after reviewing the submitted questions, it was decided that not all 120 objectives were of great importance for the course and that a higher quantity of questions about fewer objectives was preferable. As such, 75 objectives were provided for Class B and Class C and up to five students could select each objective. Students selected objectives such that the entire course was covered and all objectives were represented using materials from the course. The student choices were tracked in a shared google spreadsheet by having students put their ID numbers in the required number of cells so that almost all objectives were covered by at least one question.

Each question required the following criteria: a correct answer, three distractors, feedback explaining why the answer was correct, feedback for distractors explaining why each was incorrect (without rendering the correct answer), and identification of which level of thinking the question required based on Bloom's Taxonomy. The learning objectives for the course were assigned to present a broad and representative sample of the available materials.

The assignment was worth 15% of the final grade, marked by teaching assistants. Students were provided with a marking rubric to inform their assignment (see Table 2). Marks were based on quality of the distractors, correctness of the feedback for each distractor, correctness of the keyed answer, correct identification of each question's Bloom's Taxonomy level, and including at least two different levels of Bloom's Taxonomy in the three questions. A maximum of 14 points could be awarded per question, with three additional marks allocated to the assignment as a whole, resulting in a total score out of 45. The complete rubric has been included in Appendix B.

In Class C, the Makequiz program was updated to allow students to include graphics in their stem or possible answers. If students opted to use this feature, they were awarded a bonus mark to encourage creativity and the utilization of this new program feature.

**Table 2**

*Marking Rubric for Makequiz Assignment*

| Section | Exemplary | Satisfactory | Incomplete |
|---|---|---|---|
| Title | 1<br>Creative and comprehensive | 0.5<br>Generic, incomplete | 0<br>Missing |
| Introduction | 1<br>Provides adequate instructions about how to complete the quiz | 0.5<br>Partially complete | 0<br>Incomplete or unclear |
| Bloom's Taxonomy | 1/question<br>Question clearly relates to a specific and stated Bloom's Taxonomy objective | 0.5<br>Question partially relates | 0<br>Missing or incorrect |
| Bloom's levels | 1<br>Includes more than 1 level of Bloom's Taxonomy | 0.5<br>Only 1 level | 0<br>Missing or incorrect |
| 4 Alternatives | 1/question<br>Each question has 4 possible answers | 0.5<br>Questions have 3 options | 0<br>Less than 3 possible answers |

| Section | Exemplary | Satisfactory | Incomplete |
|---|---|---|---|
| Question Objective | 2/question<br>Question relates clearly to a specific and stated class objective | 1<br>Question partially relates to an objective | 0<br>No clear relationship between the question and the objective |
| Question Quality | 3/question<br>Clear, logical, unique, good flow, well written and appropriate level of difficulty. | 1.5<br>Clear, logical, satisfactory flow and low level of difficulty | 0<br>Unclear, poorly written and low difficulty. |
| Lecture guidelines | 2/question<br>Shows a comprehensive understanding of lecture material | 1<br>Shows a partial understanding | 0<br>Shows no understanding |
| Distractors | 2/question<br>Well thought out, appropriate level of difficulty | 1<br>Predictable, low level of difficulty | 0<br>Distractor is obvious |
| Distractor Feedback | 2/question<br>Does not give away correct answer, explains why distractor is wrong | 1<br>Partially explanation | 0<br>Incorrect, makes the correct answer obvious |
| Correct Answer Feedback | 1/question<br>Provides a clear explanation of why the answer is correct | 0.5<br>Weak explanation | 0<br>Incorrect explanation |

To further motivate students, 40 questions generated by the students were incorporated into the final exam, as literature indicates that this is an effective strategy to promote deep learning (Chickering & Gamson, 1987; Fellenz, 2004). These 40 questions accounted for 80% of the MCQs on the exam and 40% of the final exam (the other part of the exam consisted short answer questions and 10 instructor-written MCQs). The final exam accounted for 25% of students' final grades.

### *Question Quality*

The instructor and teaching assistants agreed on which questions could be considered high-quality. Selected questions were made available to all students using Makequiz which they could load from a link on the course website to allow for review prior to the final exam. Students could study questions as often as they liked and receive instant feedback with each question.

Questions were determined to be high-quality if they were clearly worded, relevant, accurate, and provided some challenge. Questions were not included online if considered too easy to guess the correct answer or potentially confusing. Student assignment scores did not determine question quality, as questions that lost marks were sometimes useful with easily modifiable errors, such as inappropriate distractors. Such errors were changed by the instructor or teaching assistant prior to dissemination to the class. Instructor-written questions were not provided to the students for study purposes.

To objectively assess the quality of student-written questions, two independent raters from the Teaching and Learning Centre were recruited to evaluate a sample of questions from Classes A and B. Both raters were experienced in teaching students and instructors how to write MCQs. One of the raters taught the initial lesson on writing MCQs to the students in this course (Classes B and C). Raters engaged in an inter-rater reliability exercise to ensure common interpretations of criteria and to clarify areas of confusion. Two questions not part of the evaluation process, were used for training the raters. After each rater scored the two questions, a discussion ensued between the raters and the instructor to discuss inconsistencies. Raters were then provided with a subset (n = 39) of randomly selected questions from Class A ($n$ = 20) and Class B ($n$ = 19). Student IDs were used for the randomization process. The first question from each students' Makequiz assignment was placed in a database and random numbers were chosen using the website: http://www.randomizer.org/form.htm. Both raters independently rated all 39 questions.

The raters used a predefined rubric as well as the list of Do's and Don'ts, which were given to the students in lecture (Table 2). The rubric was used to score the quality of students' questions (based on addressing Bloom's taxonomic levels, class objectives, quality of the question's design, and quality of feedback). The Do's and Don'ts list was used as two separate checklists to identify the presence or absence of common elements in the writing of MCQs. The Do's and Don'ts scores were summed to give a total score.

Reliability of these three scores (39 questions scored, rubric-based question quality, Do's and Don'ts) was compared between the two raters. Cronbach's alpha scores were 0.820, 0.734, and 0.951, respectively, with an inter-item correlation of 0.908 ($p < 0.001$).

## Data Analysis

Student performance was measured with six grades: Makequiz assignment, student-written final exam MCQs, instructor-written final exam MCQs, all final exam MCQs, final exam written-response, and final course grades. Correlations between Makequiz assignment scores and: a) final exam MCQ scores and b) final course grades were measured for all three semesters and overall.

Student opinions on the use of Makequiz and the assignment were collected at the end of each semester. In conjunction with university and faculty mandated evaluations, Kinesiology faculty staff unaffiliated with the course administered anonymous surveys. Students were asked to evaluate various components of the course, including Makequiz, by selecting whether they

believed the instructor should "Keep," "Drop," or "Modify" each element in future semesters. In the second semester of this study (Class B) an additional question was added to the survey asking students which components of the course they found most valuable. University-wide Universal Student Ratings of Instruction (USRI) were also used to gain an overall picture of how students viewed the course and instruction. These data helped to better understand student perceptions of Makequiz and guided the instructor in making improvements for future courses.

All analyses were conducted using SPSS v19.0 (IBM, Chicago, Illinois). Significant was set at $p \leq 0.05$.

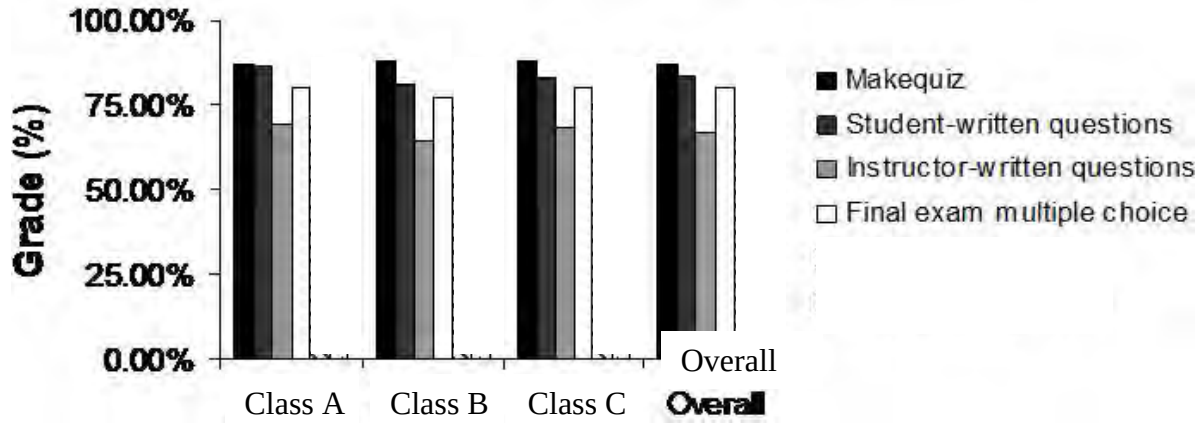
## Results

### Makequiz Grades

In Class A, 354 questions were submitted via Makequiz. Of those questions, 169 (48%) were deemed by the instructor and teaching assistants to be high enough quality for final exam preparation. In Class B, of the 333 submitted questions, 245 (74%) were provided for student study. No significant differences were noticed between the Class A and Class B questions in any of the three categories (39 questions scored, rubric-based question quality, Do's and Don'ts). In Class C, 333 questions were submitted and 176 (53%) were provided to the class as study material. It was for the Class C semester that Makequiz was modified to allow the use of graphics in stem questions and answer choices. Twenty-eight students (30 questions; 9%) took advantage of this option for extra marks on the assignment. Of these 30 questions, eight (27%) were included in the questions provided for student study. The average student grades for each semester and overall on Makequiz are displayed in Figure 3. The average student grade on Makequiz did not significantly differ between semesters (Class A: 87.2±11.1%, Class B: 87.9±11.1%, Class C: 88.0±10.8%).

**Figure 3**

*Student Grade Averages on Components of the Introduction to Research Methods Course and*




*Note.* Final Marks in Class A, Class B, Class C, and Overall.

### Student- vs. Instructor-Generated Questions

Students answered the student-written questions on the final exam correctly more often than the instructor-written questions. The average score for student-written questions was 86.5±9.1%

in Class A, 81.5±11.7% in Class B and 82.7±14.7% in Class C. The average exam score for instructor-written questions was 69.2±16.2%, 64.4±24.1% and 68.1±17.5%, respectively (Figure 3). The results were significantly different in Class A ($p$ = .014) and Class C ($p$ = .04), and in Class B there was a trend toward significance ($p$ = .053).

## Correlations

Correlations between student Makequiz grades, multiple choice scores on the final exam, and final course grades are shown in Table 3. There was a significant positive correlation for Makequiz grade and final course grade, which was strongest in Class A (Table 3).

**Table 3**

*Correlations of Makequiz Assignment Grades with Final Exam MCQ Scores and Final Course Grades for Students in the Introduction to Research Methods Course in Class A, Class B, Class C, and Overall*

| Makequiz Assignment Score | | | | |
|---|---|---|---|---|
| | Class A | Class B | Class C | Overall |
| Final Exam MCQ Score | 0.217 | 0.289 | 0.259 | 0.273 |
| Final Grade | 0.622* | 0.563* | 0.566* | 0.571* |

*Note.* *Significant correlations: $p \leq .05$.

## Student Opinions

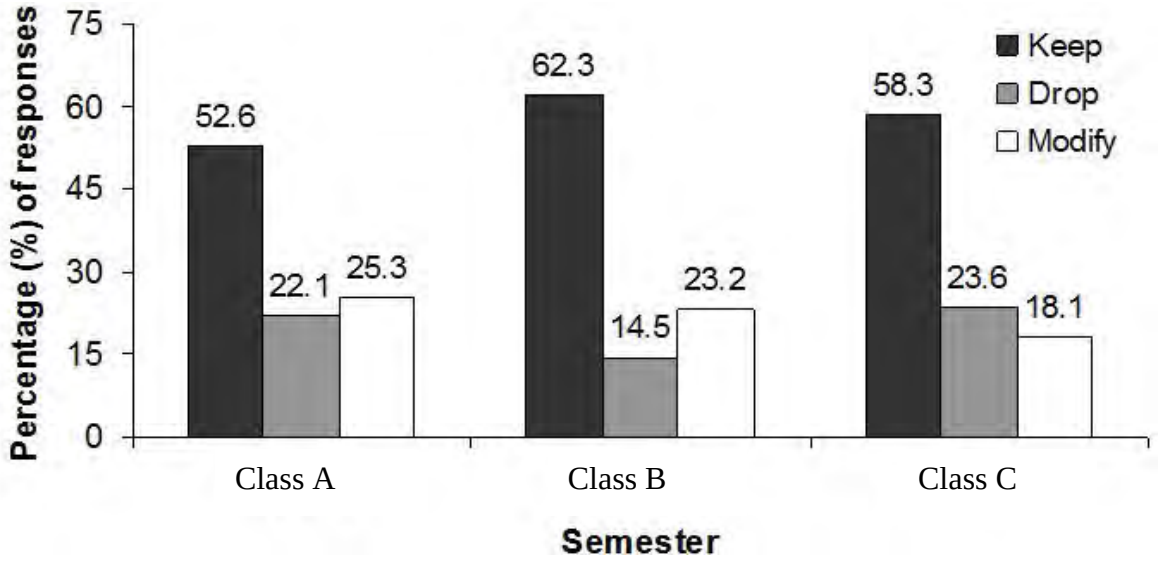
A summary of student opinions to “Keep,” “Drop,” or “Modify” Makequiz for the Introduction to Research Methods course is presented in Figure 4. A majority of students believed Makequiz should be kept in all three semesters (Class A: 52.6%, Class B: 62.3%, Class C: 58.3%), with a larger majority when including students who believe Makequiz should be kept but with modifications (Class A: 77.9%, Class B: 85.5%, Class C: 76.4%).

USRI feedback from each semester revealed that students believed this course to have a greater workload than their other courses. Of students present for the evaluation in each semester, majority ranked this course as having a “higher” or “much higher” workload (Class A: 65.0%, Class B: 71.4%, Class C: 56.6%).

In the second year of the study (Class B, Class C), 41 of 127 (32%) students who provided feedback on the course evaluation named Makequiz as the, or one of the, most valuable components of the entire course.

**Figure 4**

*Student Opinions about Whether to Keep, Drop, or Modify Makequiz as a Component of the Introduction to Research Methods Course in Class A, Class B, and Class C*




## Discussion

The present study finds support for Makequiz as a study tool, as the majority of students reported a preference to keep and/or modify this component of the course rather than drop it, and approximately one third of students who were asked about components of the course stated it was the most valuable assignment.

Considering the results, it should be remembered that student opinions may have been affected by factors beyond the control of the instructor. For example, some students encountered technical difficulties on their personal computers during program use, such as inability to connect to the network. Additionally, the number of assignments in the course overall was identified as being much higher than other courses. Furthermore, USRI course evaluations were done near the end of classes for the semester, shortly after receiving their Makequiz assignment grades but prior to writing the final exam (and likely prior to student preparation for the final).

This could suggest that students had not yet decided on the usefulness of the assignment in the grand scheme of the course, and particularly in relation to studying for the final exam. However, it is likely not practical to ask students to complete a survey after the course has finished, as this would result in an extremely low response rate.

## Student Performance

### *Student vs. Instructor Questions*

Generally stated, students performed better on student-written questions than instructor-written questions. This may be because students recognized questions and answers from those provided to them prior to the final exam. However, memorization seems unlikely as between 169-245 questions were provided to students for study purposes. Familiarity with the questions, rather than memorization of all questions, is likely and would require further investigation and data

collection to make clear. The discrepancy between student-written and instructor-written question performance is likely not due to worse quality of student-written than instructor-written questions because only high-quality questions were included on the exam.

Additionally, McLeod and Snell (1996) observed no difference in quality between student- and faculty-written questions. Though quality may be comparable, difficulty may be higher with instructor-written versus student-written questions due to the instructor's deeper understanding of the course concepts. Previous research has shown that students are more likely to critically examine peer-created questions than instructor-created material when studying (Sanchez-Elez et al., 2014). Further investigation is needed to determine specific reasons for the difference in student success on student- and instructor-written questions.

### *Makequiz Course-Grade Correlation*

There was a significant positive correlation between Makequiz grade and final course grade. This was strongest in Class A (Table 3). Though causation cannot be determined in this study, one possibility is general student effort in school.

Eager students who aim for high grades likely put in strong efforts on assignments to earn desired grades. This agrees with Green (2016) who reported that student scores positively increased with the use of SRS in class, but that this was also highly correlated with SAT scores; students with higher SAT scores received more benefit from the SRS in class.

Another possibility is that working hard to do well on the Makequiz assignment potentially guided students to deeper learning of course material, facilitating higher grades in the course. Further research is required to discover causation of such correlations.

## Learning Benefits

### *Active Studying*

Taking an active role in learning through writing MCQs on course objectives can lead to better grades (Sanchez-Elez et al., 2014). As mentioned, justifying answers during studies improves test scores (Koretsky et al., 2016). Focusing on higher levels of Bloom's Taxonomy, the assignment outlined in this study gave students opportunity for higher-order learning. Although considering potential confusions and writing effective feedback, students were required to view their questions from multiple perspectives and understand course material on a deeper level. Whether learning was impacted by this strategy remains to be investigated.

### *Metacognition*

Reviewing peer questions provided an opportunity to deepen understanding of course objectives that were not selected by students themselves for the Makequiz assignment. As a function of the Makequiz program, feedback was provided after achieving a correct answer and as an explanation for an incorrect answer choice. Such feedback is rarely available to students although preparing for exams. This accurate assessment of performance, which is generally poorly estimated by students themselves (Pressley & Ghatala, 1990), may have helped students prepare for the final exam. In other words, metacognition may have improved through receiving accurate insight into the student proximal zone of knowledge (Metcalfe, 2009). Providing immediate

feedback with MCQs although preparing for an exam could, therefore, be a means for helping learners develop metacognitive judgement and use metacognitive tools (Huff & Nietfeld, 2009).

Alternatively, students with greater metacognitive skill may be more likely to use strategies such as reviewing peer-created questions in preparation for exams, which can result in improvement between pre-and post-tests (de Carvalho Filho, 2010; Schraw, 1994).

Further research is required to investigate the nature of such a relationship between metacognition and test performance with Makequiz. A possible approach for investigation is to examine the rate at which metacognitive skills develop, to measure if feedback through Makequiz has the potential to meaningfully impact metacognition, or if it differentially helps students with varying metacognitive abilities.

### *MCQs*

There are other potential, though currently immeasurable, benefits to the students from the focus on MCQ development. This course contained a large number of first- and second-year students, so the instruction on how to write high-quality MCQs and recognize poor questions may assist them in test-taking for the remainder of their undergraduate careers. Although no empirical system was employed to determine the quality of questions on these exams between semesters, an attempt was made to choose high-quality questions from students who employed a higher level of Bloom's taxonomy. Improving the quality of the questions could increase student success and the reliability of the questions although also reinforcing student learning and potentially decreasing test anxiety (Bacon, 2003; Cizek & O'Day, 1994; Foos, 1989; Foos et al., 1994; Green, 1997). Further research is needed to measure the impact of Makequiz on anxiety levels and impacts on learning.

### *Peerwise*

Peerwise is an online learning tool, similar to Makequiz. It provides a forum for student-generated questions, with opportunity for rating question difficulty and quality, and peer discussions through commenting and responding (https://peerwise.cs.auckland.ac.nz/). Positive effects of using Peerwise have been reported (Denny et al., 2008a; Draper, 2009; Duret et al., 2018; McKenzie & Roodenburg, 2017; Sykes, Hamer, & Purchase, 2018). Students enjoy it (Denny et al., 2008b; Denny et al., 2009; Rhodes, 2015) and perceive a learning benefit relative to writing online quizzes (McKenzie & Roodenburg, 2017). Observations of academic success have been inconsistent with some learning improvement with PeerWise (Denny et al., 2008; Duret et al., 2018), and some with no differences in recall or exam performance (McKenzie & Roodenburg, 2017; Rhodes, 2015). There is a consensus in the helpfulness of peer discussions (McKenzie & Roodenburg, 2017; Sykes et al., 2018) with positive correlations of grades with commenting activity (Denny et al., 2008b; Denny et al., 2008c; Duret et al., 2018; Rhodes, 2015).

Sykes and colleagues (2018, p. 2) labels learning enhancement from Peerwise as "uncertainty-resolution." The opportunity for uncertainty that Peerwise provides is important for learning because, they argue, learning deepens as students work to resolve uncertainties from peer-written questions (Draper, 2009; Sykes et al., 2018). Discussion to resolve uncertainty with peers helps students gain fuller grasp of material (Draper, 2009; Sykes et al., 2018). Another enhancement of learning from Peerwise is said to be creating distractors for questions because a

good grasp of material is needed for writing good distractors (Denny et al., 2008b). Though the programs are similar, there are differences (Table 4).

**Table 4**
*Comparison of Makequiz and Peerwise functions.*

| | Makequiz | Peerwise |
|---|---|---|
| Commenting and discussing | No | Yes – all discussion appears after answering a question and option to comment |
| Feedback | Yes – Immediate, specific to answer choice | Yes – After answering each question, feedback written for the question and answers together, shows prevalence of other student answers |
| Requires internet | No – generates Html5 file | Yes |
| Compiles student answers | No – must do it manually | Yes – automatically incorporates all student answers together online |

When a student picks an answer in Makequiz, feedback is immediately given explaining why that answer is or is not correct. Students then have the resources to reconsider and rethink before answering again. In Peerwise, feedback is given for all answer choices at the same time after answering. Peerwise allows students to discuss and provide feedback, but Makequiz does not. Peerwise is online and Makquiz generates an html, meaning it opens in an internet browser but does not require an internet connection. The offline aspect of Makequiz requires the instructor to combine all student questions together manually, and Peerwise combines them together online. More research is needed to measure learning impacts from Makequiz, but based on Peerwise research and its similarities to Makequiz, Makequiz holds potential to improve learning. Additionally, the positive perceptions of Makequiz provide support for investigation of Peerwise in research and statistics courses, which has not yet been examined.

The instructor of the present study used Peerwise in a subsequent and similar fashion to the Makequiz assignment. All students were taught how to write good quality MCQs with a focus on high-order learning, and were required to submit three questions on three different course objectives. To encourage engagement, students were required to comment on three different peer-written questions addressing any of: question quality, content, answer options, explanation, grammar/spelling. It would be valuable to compare learning with these different programs to understand which aspects are helpful for students and to know how to maximize student learning and success.

## Co-creation, Constructivism and Relationships

Dollinger et al. (2018) report that students are increasingly interested in working collaboratively and interacting with each other and as partners with educators in the co-creation of materials, contexts, assessments, and curricular opportunities. This is inherently a constructivist practice involving the generation and modification of knowledge and skills based on the modification of internal representations and models through interactions with the environment

(Jonassen, 1994). "Deep and active collaboration between students and teachers is vital for successful co-creation of the learning experience (Doyle & Buckley, 2022, p. 1727). In this study, Makequiz provided the context necessary for this collaborative co-creation to function and for a constructivist paradigm to be embraced. This may be important for another reason, namely that of feeling connected to the instructor, to other students and to the course material.

Roy Baumeister and Mark Leary (1995) state that:

> People form social attachments readily under most conditions and resist the dissolution of existing bonds. Belongingness appears to have multiple and strong effects on emotional patterns and on cognitive processes. Lack of attachments is linked to a variety of ill effects on health, adjustment, and well-being…Existing evidence supports the hypothesis that the need to belong is a powerful, fundamental, and extremely pervasive motivation. (p. 497)

In the context of Makequiz, we can see that there is an element of belonging and relationship building that occurs at a subtle level by allowing students to co-create the space for assessment and feedback with guidance from the instructor. Sidorkin (2023) states, "Learning does not happen between a student and a book; it happens in the space between a student and a teacher (or a peer). Learning happens when they solve a problem together (p. 59). In the Makequiz situation there are multiple problems being solved: that of the students completing an assignment, studying for a test, predicting possible incorrect answers, and creating feedback; that of the instructor guiding students, verifying questions, and creating assessments; and those in the common space of the co-generation of a relationship that recognizes the different requirements and tasks that feed the teaching and learning processes. In the wake of Covid isolations, it is this last factor, the co-construction of common spaces of mutual understanding with differing roles and responsibilities that is perhaps the most germane and impactful.

## Conclusions

Students were taught to write high-quality MCQs and generally enjoyed using Makequiz. Having students write course-related MCQs for the Makequiz assignment was positive for many students and may continue to have beneficial effects for them. Using Makequiz allows many students to think critically about course material, and to fully grasp concepts through understanding content and sources of confusion. Makequiz provides students with instant, detailed feedback on every answer given, generally resulting in a stronger performance on student-written questions on the final exam. Based on student feedback, Makequiz is an appreciated study tool.

Makequiz targets higher-order learning but its effect on academic success remains unclear. Further investigation is necessary to measure relative anxiety levels with and without Makequiz, the development of metacognitive skills during this process for different types of students, and to determine sources of differences in success on student- versus instructor-written questions. Makequiz should be compared to PeerWise to maximize learning with MCQs.

Specific changes in student investment in the course as well as test performance because of Makequiz are yet to be measured. With Makequiz students generate questions and answers, including distractors, *anticipating* peer errors and misconceptions. This is a different skill from the typical collaboration as it requires taking and understanding the role of another student.

PeerWise uses direct collaborative methods to improve upon questions as they are being created, possibly making for a more effective experience with regards to feedback. Both programs

and methods provide students with ways to learn objectives, create and understand good distractors, and give quality feedback that improves learning. Although they have their differences, each program has been developed with similar intentions, namely to improve student learning through collaboration and the use of multiple-choice questions.

## References


Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short-answer questions in a marketing context. *Journal of Marketing Education, 25*(1), 31–36. https://doi.org/10.1177/0273475302250570

Bates, S. P., Galloway, R. K., Riise, J., & Homer, D. (2014). Assessing the quality of a student-generated question repository. *Physical Review Special Topics-Physics Education Research, 10*(2), 020105. https://doi.org/10.1103/PhysRevSTPER.10.020105

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497-529. https://doi.org/10.1037/0033-2909.117.3.497

Berry, J. W., & Chew, S. L. (2008). Improving learning through interventions of student- generated questions and concept maps. *Teaching of Psychology, 35*(4), 305–312. https://doi.org/10.1080/00986280802373841

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longmans.

Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning, 11*(2), 215–235. https://doi.org/10.1007/s11409-015-9142-6

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin, 39*(7), 2-6. https://aahea.org/articles/sevenprinciples1987.htm

Cizek, G. J., & O'Day, D. M. (1994). Further investigation of nonfunctioning options in multiple-choice test items. *Educational and Psychological Measurement, 54*(4), 861–872. https://doi.org/10.1177/0013164494054004002

Craft, J. A., Christensen, M., Shaw, N., & Bakon, S. (2017). Nursing students collaborating to develop multiple-choice exam revision questions: A student engagement study. *Nurse Education Today, 59*, 6–11. https://doi.org/10.1016/j.nedt.2017.08.009

de Carvalho Filho, M. K. (2010). Assessing changes in performance and monitoring processes in individual and collaborative tests according to students' metacognitive skills. *European Journal of Cognitive Psychology, 22*(7), 1107–1136. https://doi.org/10.1080/09541440903336555

Denny, P., Luxton-Reilly, A., & Hamer, J. (2008a). The PeerWise system of student contributed assessment questions. In *Proceedings of the Tenth Conference on Australasian Computing Education* (pp. 69–74).

Denny, P., Luxton-Reilly, A., & Hamer, J. (2008b). Student use of the PeerWise system. *Association for Computing Machinery Special Interest Group on Computer Science Education Bulletin, 40*(3), 73–77. http://doi.org/10.1145/1597849.1384293

Denny, P., Hamer, J., Luxton-Reilly, A., Purchase, H. (2008c). PeerWise: Students sharing their multiple choice questions. In *Proceedings of the Fourth International Workshop on Computing Education Research* (pp. 51–58). Association for Computing Machinery. https://doi.org/10.1145/1404520.1404526

Denny, P., Luxton-Reilly, A., Hamer, J., & Purchase, H. (2009). Coverage of course topics in a student generated MCQ repository. *Association for Computing Machinery Special Interest Group on Computer Science Education Bulletin*, *41*(3), 11-15. https://doi.org/10.1145/1595496.1562888

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning*, *2*(2), 4. https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Ding, L., Reay, N. W., Lee, A., & Bao, L. (2009). Are we asking the right questions? Validating clicker question sequences by student interviews. *American Journal of Physics, 77*(7), 643-650. http://doi.org/10.1119/1.3116093

Dollinger, M., Lodge, J., & Coates, H. (2018). Co-creation in higher education: Towards a conceptual model. *Journal of Marketing for Higher Education, 28*(2), 210–231. https://doi.org/10.1080/08841241.2018.1466756

Doyle, E., & Buckley, P. (2022). The impact of co-creation: An analysis of the effectiveness of student authored multiple choice questions on achievement of learning outcomes, *Interactive Learning Environments*, *30*(9), 1726–1735. https://doi.org/10.1080/10494820.2020.1777166

Draper, S. W. (2009). Catalytic assessment: Understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology, 40*(2), 285–293. https://doi.org/10.1111/j.1467-8535.2008.00920.x

Duret, D., Christley, R., Denny, P., & Senior, A. (2018). Collaborative learning with PeerWise. *Research in Learning Technology*, *26*. https://doi.org/10.25304/rlt.v26.1979

Egelandsdal, K., & Krumsvik, R. J. (2017). Clickers and formative feedback at university lectures. *Education and Information Technologies*, *22*(1), 55–74. https://doi.org/10.1007/s10639-015-9437-x

Fellenz, M. R. (2004). Using assessment to support higher level learning: The multiple choice item development assignment. *Assessment & Evaluation in Higher Education, 29*(6), 703-719. https://doi.org/10.1080/0260293042000227245

Flavell, J., & Wellman, H. (1977). Perspectives on the development of memory and cognition. In J. R. V. Kail, & J. Hagen (Eds.), *Metamemory* (pp. 3-33). Erlbaum.

Foos, P. W. (1989). Effects of student-written questions on student test performance. *Teaching of Psychology*, *16*(2), 77–78. https://doi.org/10.1207/s15328023top1602_10

Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology*, *86*(4), 567-576. https://doi.org/10.1037/0022-0663.86.4.567

Foote, C. J. (1998). Student-generated higher order questioning as a study strategy. *The Journal of Educational Research*, *92*(2), 107–113. https://doi.org/10.1080/00220679809597582

Foster, C. (2011). Student-generated questions in mathematics teaching. *Mathematics Teacher, 105*(1), 26–31. https://doi.org/10.5951/mathteacher.105.1.0026

Gonzalez-Cabezas, C., Anderson, O. S., Wright, M. C., & Fontana, M. (2015). Association between dental student-developed exam questions and learning at higher cognitive levels. *Journal of Dental Education*, *79*(11), 1295–1304. https://doi.org/10.1002/j.0022-0337.2015.79.11.tb06025.x

Gooi, A. C., & Sommerfeld, C. S. (2015). Medical school 2.0: How we developed a student-generated question bank using small group learning. *Medical Teacher, 37*(10), 892–896. https://doi.org/10.3109/0142159X.2014.970624

Green, A. (2016). Significant returns in engagement and performance with a free teaching app. *Journal of Economics Education, 47*(1), 1–10. https://doi.org/10.1080/00220485.2015.1106359

Green, D. H. (1997). Student-generated exams: Testing and learning. *Journal of Marketing Education, 19*(2), 43–53. https://doi.org/10.1177/027347539701900205

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning, 4*(2), 161–76. https://doi.org/10.1007/s11409-009-9042-8

Jonassen, D. H. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology, 34*(4), 34–37.

Kay, A. E., Hardy, J., & Galloway, R. K. (2020). Student use of PeerWise: A multi-institutional, multidisciplinary evaluation. *British Journal of Educational Technology, 51*(1), 23–35. https://doi.org/10.1111/bjet.12754

Kim, M.-K., Patel, R. A., Uchizono, J. A., & Beck, L. (2012). Incorporation of Bloom's Taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. *American Journal of Pharmaceutical Education, 76*(6), 114. https://doi.org/10.5688/ajpe766114

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*(1), 111–126. https://doi.org/10.1207/s15326985ep2701_8

Koretsky, M. D., Brooks, B. J., White, R. M., & Bowen, A. S. (2016). Querying the questions: Student responses and reasoning in an active learning class. *Journal of Engineering Education, 105*(2), 219–244. https://doi.org/10.1002/jee.20116

Lord, T., & Baviskar, S. (2007). Moving students from information recitation to information understanding - exploiting Bloom's Taxonomy in creating science questions. *Journal of College Science Teaching, 36*(5), 40.

Marinagi, C., Skourlas, C., & Belsis, P. (2013). Employing ubiquitous computing devices and technologies in the higher education classroom of the future. *Procedia-Social and Behavioral Sciences, 73*, 487–494. https://doi.org/10.1016/j.sbspro.2013.02.081

Mavis, B. E., Cole, B. L., & Hoppe, R. B. (2001). A survey of student assessment in US medical schools: The balance of breadth versus fidelity. *Teaching and Learning in Medicine, 13*(2), 74–79. https://doi.org/10.1207/S15328015TLM1302_1

McKenzie, W., & Roodenburg, J. (2017). Using PeerWise to develop a contributing student pedagogy for postgraduate psychology. *Australian Journal of Educational Technology, 33*(1). https://doi.org/10.14742/ajet.3169.

McLeod, P., & Snell, L. (1996). Student-generated MCQs. *Medical Teacher, 18*(1), 23–25. https://doi.org/10.3109/01421599609040257

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159–163. https://doi.org/10.1111/j.1467-8721.2009.01628.x

Morton, D. A., & Colbert-Getz, J. M. (2017). Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's Taxonomy. *Anatomical Sciences Education, 10*(2), 170–175. https://doi.org/10.1002/ase.1635

Parkes, J., & Zimmaro, D. (2016). *Learning and assessing with multiple-choice questions in college classrooms* (1st ed.). Routledge. https://doi.org/10.4324/9781315727769

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25*(1), 19-33. https://doi.org/10.1207/s15326985ep2501_3

Rhodes, J. (2015). Using PeerWise in nursing education – a replicated quantitative descriptive research study. *Kai Tiaki Nursing Research, 6*(1), 10–15.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181–221. https://doi.org/10.3102/00346543066002181

Sanchez-Elez, M., Pardines, I., Garcia, P., Miñana, G., Roman, S., Sanchez, M., & Risco, J. L. (2014). Enhancing students' learning process through self-generated tests. *Journal of Science Education and Technology, 23*(1), 15–25. https://doi.org/10.1007/s10956-013-9447-7

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology, 19*(2), 143–154. https://doi.org/10.1006/ceps.1994.1013

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Sidorkin, A. (2023). *Pedagogy of relations*. Routledge. https://doi.org/10.4324/9781003179627

Song, D., Oh, E. Y., & Glazewski, K. (2017). Student-generated questioning activity in second language courses using a customized personal response system: A case study. *Educational Technology Research and Development, 65*(6), 1425–1449. https://doi.org/10.1007/s11423-017-9520-7

Stanley, T. (2021). *Performance-based assessment for 21st-century skills*. Routledge. https://doi.org/10.4324/9781003237129

Sykes, A., Hamer, J., & Purchase, H. (2018). PeerWise as a vehicle for uncertainty-resolution to enhance student learning. *Journal of Perspectives in Applied Academic Practice, 6*(1), 55–65. https://doi.org/10.14297/jpaap.v6i1.317

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC medical education, 9*(1), 40. https://doi.org/10.1186/1472-6920-9-40

Thompson, A. R., & O'Loughlin, V. D. (2015). The Blooming anatomy tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anatomical Sciences Education, 8*(6), 493–501. https://doi.org/10.1002/ase.1507

Walsh, C. M., & Seldomridge, L. A. (2006). Critical thinking: Back to square two. *The Journal of Nursing Education, 45*(6), 212–219. https://doi.org/10.3928/01484834-20060601-05

Ware, J., & Vik, T. (2009). Quality assurance of item writing: During the introduction of multiple choice questions in medicine for high stakes examinations. *Medical Teacher, 31*(3), 238–243. https://doi.org/10.1080/01421590802155597

Zaidi, N. B., Hwang, C., Scott, S., Stallard, S., Purkiss, J., & Hortsch, M. (2017). Climbing Bloom's taxonomy pyramid: Lessons from a graduate histology course. *Anatomical Sciences Education, 10*(5), 456-464. https://doi.org/10.1002/ase.1685

Zaidi, N., Grob, K., Monrad, S., Kurtz, J. , Tai, A., Ahmed, A., Gruppen, L. & Santen, S. (2018). Pushing critical thinking skills with multiple-choice questions: Does Bloom's Taxonomy work? *Academic Medicine*, *93*(6), 856–859. https://doi.org/10.1097/ACM.0000000000002087

Zimmaro, D. (2004, August 19). *Writing good multiple choice exams.*University of Texas at Austin. https://pmm.uinsu.ac.id/wp-content/uploads/2021/07/Panduan-membuat-butir-soal-pilihan-ganda-HOT.pdf

# Appendix A
## Lecture on Creating High-Quality Multiple-Choice Questions

**Kinesiology 213**
**Introduction to Research in Kinesiology**

Multiple Choice, Make Quiz and Distractors

### Multiple Choice Questions

Multiple Choice Questions consist of:

- The question is called the 'stem'
- The choices from which the student chooses are called 'the choices'
- The choice that is correct is called, 'the correct answer (or best choice)'
- The incorrect choices are called the distractors

### Picking Distractors for Multiple Choice Exams

Distractors are the incorrect answers in a multiple choice question. Poorly chosen distractors can make your multiple choice questions much easier or harder than intended.

- Distractors should highlight common student mistakes
- Only have one correct answer
- Avoid always/never and all/none of the above
- Four is a good number but only if you have quality choices
- Distractors should be put in a logical order whenever such an order exists. For example, numeric answers should be listed in increasing or decreasing order.
- Stem and distractors should be grammatically consistent
- Question difficulty is controlled through the distractors
- Avoid complex choices
- How do you know if your question is a good question?

### Definitely Do!

1. Put all important information into the stem, options don't add information
2. Ensure the stem has all of the info needed to be clear and specific
3. Ensure you pose a direct, unambiguous question
4. Ensure there really is one best answer – not debatable
5. Ensure answers are not dependent upon unstated conditions
6. Write questions based on significant content or skill based objectives

Kinesiok

### Avoid Like the Plague

1. Don't aim to be ambiguous in your writing - use external ambiguity not internal ambiguity (More in a minute)
2. Avoid negatives in your stem questions like "Which of these are NOT..."
3. Avoid questions that aim to deliberately trick people, e.g. "Some months have 31 days, how many have 28?"
4. Get rid of distracters that are not plausible, or open the stem to alternate interpretations.
5. Make sure answers (such as time interval and dates) do not overlap. This can be confusing.
6. Avoid true answers that do not answer the question.

Kinesiok

## Evaluation Questions

What would an evaluation question ask you to do?

- Consider the actions taken by others to judge their decision making/ performance.
- Determining the best course of action, given the criteria
- Judging the validity of conclusions

Kinesiology

## Cuing No-Nos in MC Questions

Clang Associations:
Show up when there is an answer in the stem.

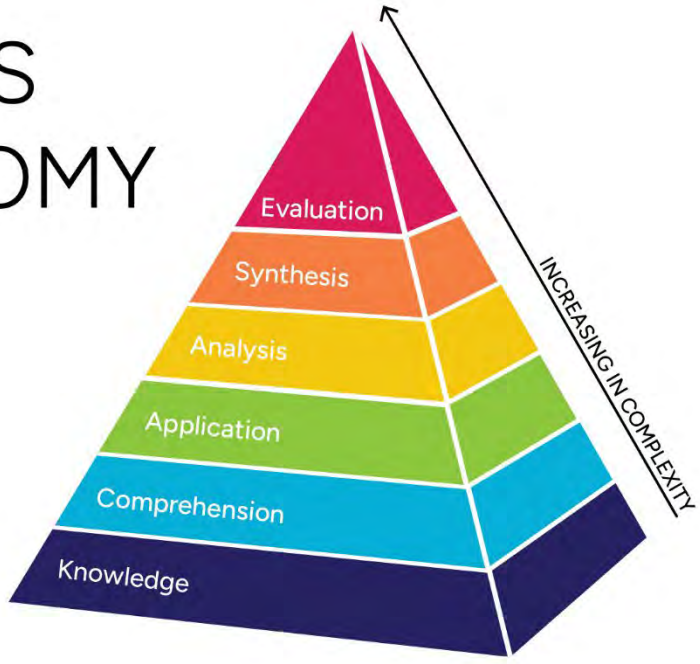
For Example:

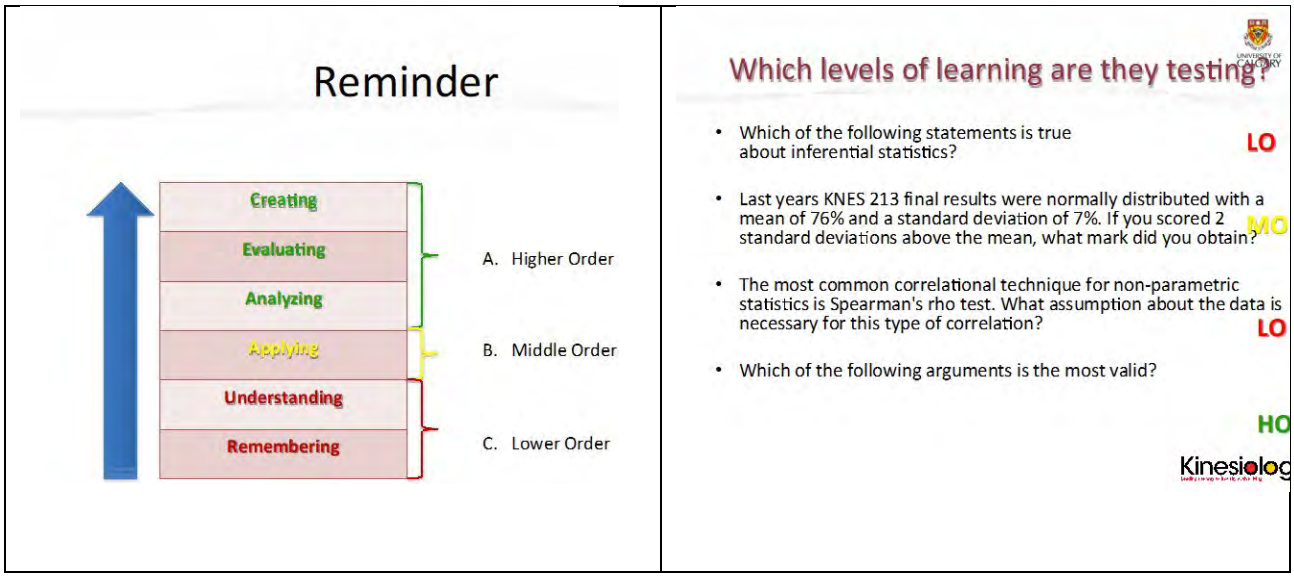
Why was Germany's WWI invasion of France through Belgium ill considered?

a. The French fought fiercely, fending off the assault
b. By invading through Belgium Germany provoked the British
c. The German military could have attacked France directly

Kinesiology

## BLOOMS TAXONOMY




Reminder
Creating
Evaluating
Analyzing
Applying
Understanding
Remembering
A. Higher Order
B. Middle Order
C. Lower Order
Which levels of learning are they testing?
Which of the following statements is true about inferential statistics?
LO
Last years KNES 213 final results were normally distributed with a mean of 76% and a standard deviation of 7%. If you scored 2 standard deviations above the mean, what mark did you obtain?
MO
The most common correlational technique for non-parametric statistics is Spearman's rho test. What assumption about the data is necessary for this type of correlation?
LO
Which of the following arguments is the most valid?
HO
Kinesiolog

## Appendix B
## Marking Guide

Marking Guide – used to summarize scores and to determine if a question should be included for other students to use for study purposes.

| Topic | Value | Mark | Good Quality 'Keep' |
|---|---|---|---|
| Title | 1 | | |
| Introduction | 1 | | |
| Identify Objective Bloom's Taxonomy | 3 | | |
| Two different Levels of Bloom Taxonomy | 1 | | |
| Four Alternatives | 3 | | |
| Question Objective | 6 | | |
| Question Quality | 9 | | |
| Incorporate guidelines from lecture | 6 | | |
| Appropriate Distractors | 6 | | |
| Appropriate feedback for Distractors | 6 | | |
| Appropriate Feedback for Correct Answer | 3 | | |
| Total | 45 | | |