Advances in Engineering Education

2023: VOLUME 11 ISSUE 2
DOI: 10.18260/3-1-1153-36043

Create-Rank-Compete Crowdlearning

COURTNEY J. BURRIS

ALEXANDER NIKOLAEV

HIMANGSHU PAUL

AND

LAN PENG
Department of Industrial and Systems Engineering
University at Buffalo, Buffalo, NY

ABSTRACT

Create-Rank-Compete (CRC) is a teaching method designed to promote learner-to-learner, learner-to-material, and learner-to-instructor engagement through asynchronous and synchronous online interactions. The CRC method is rooted in the constructivism theory of learning; by design, it is a Bruner's spiral that supports metacognitive skill development through problem posing and peer assessment activities. The method is implementable in both seated and remote courses. Periodically during a course, students are asked to create multiple choice questions based on the covered material; then they perform pairwise comparisons of the de-identified peers' creations; from the ranked list of the questions, the instructor is able to track the class progress and use this student-generated content to compose quiz materials for testing and discussion. This paper details the CRC method and reports on its first implementation, which took place during the outbreak of the Covid-19 pandemic, in the semester when a large public university in the northeastern United States switched from live to online teaching. Students engaged with the CRC activities two times more than required, and reported high enjoyment and knowledge acquisition value delivered by the method.

Key words: Active Learning, Engagement, Peer Review

INTRODUCTION

Online learning communities are changing with the advance of technology, growing number of students, and, as of late, the Covid-19 pandemic. This produces a need for more intelligent, interactive systems that can handle massive learning needs whilst ensuring quality and reliability within the system.

In March of 2020 many colleges and universities were forced to close their doors for more than a year and switch to distance learning due to the Covid-19 pandemic. Synchronous and asynchronous teaching methods were adopted at large in a haste so as not to halt the courses which were already in full swing. One popular tool for synchronous learning is Zoom videoconferencing: it features a virtual board, screen sharing and breakout rooms capabilities, which allow students and instructors to communicate and engage. However, an upturn in "Zoom fatigue" signals that video conferencing tools are becoming taxing to students (Morris 2020). This brings to the forefront the challenges to overcome for asynchronous learning to be successful: (1) the need to promote engagement, especially in the mixed offline/online mode of learning, and (2) the need for creative practices to spark just-in-time feedback exchange between the students and instructor. There is a high demand for effective online teaching methods — to be put to practice in a stand-alone format or become an engagement-promoting course component — which can be implemented in small and large virtual class settings.

We present Create-Rank-Compete (CRC) — a teaching method comprising three phases. The CRC question creation phase prompts students to generate questions pertaining to the recent class' material, increasing their engagement with this material. The CRC question ranking phase tasks students with conducting pairwise comparisons of peers' questions, increasing both their engagement with material and each other. The CRC quiz competition phase uses a gamified quiz that ranks participants by performance to provide students with formative feedback on their understanding of course material whilst facilitating a structured live review of said material. The key premise of CRC is to use alternative assessments in the form of question creation, and peer feedback in the form of question ranking, to promote engagement within a course. Indeed, alternative assessments and the use of peer feedback have been shown to promote a deeper level of comprehension than instructor-focused lectures (Cannella, Ciancimino, and Campos 2010; Ciancimino and Cannella 2008; Sadler 1989).

The CRC method is an iterative process situated in social constructivism (Aalst 2009) where students' conceptual knowledge is constructed through not only course instruction but through interactions with the instructor and their peers. Student-generated questions inform the instructor on knowledge gaps and on how best to spend class time. In turn, the questions that the instructor selects for the live quizzes/discussions (done in-class, simultaneously for all students) inform the students on which pieces of the subject matter are the most important to know and retain. This creates a learner-focused instructional environment based upon identifiable needs, whilst pacing the instruction in a way that takes into account student capability and progress.

The CRC method is derived from the seminal research into Crowdlearning, conducted for several years in an engineering department at a large public university (Farasat et al. 2017). The first CRC implementation took place in the Spring 2020 semester, prior to transitioning to distance learning due to Covid-19. As the course switched to the fully online instruction, the CRC method remained unchanged. Experiences from this implementation motivated refinements; the refinement cycles produced the final CRC framework presented here.

This paper is structured as follows. Section "Theoretical Bases of CRC" provides the theoretical foundation for the CRC methodology. Section "Methodology" provides an explanation of the purpose, route to implementation, and grading scheme, as well the algorithms utilized within the CRC process. Section "Results and Takeaways" reports on the observations from the initial CRC implementation and summarizes the lessons learned, students perceptions of the method, and the impact the method has on students' grades. Section "Conclusion" completes the paper and discusses immediate and future steps for further Crowdlearning-inspired developments. The Appendix details data handling algorithms that facilitate the implementation of CRC in practice.

THEORETICAL BASES OF CREATE-RANK-COMPETE

This section covers the theoretical bases for the CRC method, with a focus on three theories: problem-posing, peer assessment as a learning tool, and metacognitive development through gamification.

Curricular design has historically been rooted in basic science models, rather than learning models; this is evidenced by a scarcity of research-based principles used within higher education, specifically in the Science Technology Engineering Math (STEM) fields (Braha and Maimon 1997). To overcome this, learner engagement has become a focal point of new instructional approaches. Activities which incorporate problem-posing, peer assessment, and gamification have been shown to engage students in the process of active learning, which encourages the development of students' metacognitive skills and helps to achieve desired learning outcomes (see (Brindley and Scoffield 1998; English 1998; Li and Tsai 2013; Purchase 2000; K. J. Topping and Ehly 2001; Veenman 2012)).

The CRC method is situated in the constructivist theory of learning wherein the main goal of instruction is in knowledge construction and cognitive development rather than acquisition of specific behaviors and skills (Fosnot and Perry 1996). The CRC process structure follows a compressed Bruner's spiral curriculum. Bruner's sprial curriculum was originally conceived with the goal of improving cognitive development of children, however, this theory can be expanded and applied to a wider range of student groups (Masters and Gibbs 2007; Johnston 2012). The three key ideas behind a Bruner's spiral cycle are reinforcement of topics, progression of complexity, and connection to prior knowledge. Thus, concepts are reinforced by revisiting the same topic multiple times throughout a student's school career, or in the compressed case, throughout a single semester (Gibbs 2014). Each time a topic is revisited, the complexity increases, as cognitive development progresses. The connection between topics ensures that the students access prior knowledge, which is a pillar of constructivism.

The fundamental idea behind the CRC method lies in crowdsourced problem-posing. Problem-posing is defined as the process of creating a new problem/question or reformulating given problems based on conceptual content (Mishra and Iyer 2015; Silver et al. 1996). Although not typically adopted in conventional education, the **theory of problem-posing** indicates that learners reap both motivational and cognitive benefits as a result of engaging in problem-posing activities (Rotherham and Willingham 2009; Shute and Torres 2012). Further, active participation in problem-posing cultivates an inquisitive culture within the classroom, i.e., students are more apt to and capable of asking questions independent from the problem-posing activity (Chin 2002; Musingafi, Muranda, et al. 2014). Unfortunately, although user-contributed tools have been extended remarkably in students' social entertainment, there is little effort toward bringing user content-creation into the classroom (Beal, Cohen, et al. 2012).

Peer assessment as a learning tool is the second foundational theory behind the CRC method. Peer assessment refers to an activity in which students evaluate and provide feedback to each other. Using peer assessment as a learning tool is advantageous to the teaching-learning process as it increases students' intrinsic motivation to engage in active learning, positively affects their self-efficacy, and encourages student accountability (Palmero and Rodríguez 2012). Students benefit from peer assessment both when receiving feedback and when providing feedback to their fellow students. There is evidence supporting the claim that students are more receptive to peer feedback than to instructor feedback, further, this creates an opportunity for students to reflect upon and revise their work which hones their domain-specific skills (Van Zundert, Sluijsmans, and Van Merriënboer 2010). Assessors have a role in reviewing, giving feedback, identifying misconceptions, and considering deviations from the ideal (Van Lehn et al. 1995), which, in turn, helps them to consolidate, reinforce, and deepen understanding of subject matter as well as developing specific skills (K. Topping 1998).

The last component of our theoretical framework is **metacognition, with a focus on game-based principles**, i.e., gamification. Metacognition is the "knowledge, awareness, and control of one's own learning" (Gunstone and Mitchell 2005) and is heavily intertwined with conceptual change. Conceptual change occurs when students receive new information that provokes the development of fundamentally new concepts via a restructuring of preexisting knowledge components (Disessa 2002; Linn, Clark, and Slotta 2003). However, recognizing, evaluating, and structuring conceptions, new and preexisting, requires metacognitive awareness, which many students lack (Bransford et al. 1999). Two ways to encourage metacognition in classroom are outlined as follows. First, metacognition can be stimulated through the use of game-based principles as this introduces competition into the classroom; competition is presumed to energize individuals, however, its effect on a learner depends on how they channel this energy – toward performance-approach or performance-avoidance (Elliot 2020). Competition is known to promote normative competence evaluation and activate social comparison processes (Garcia, Tor, and Schiff 2013). Second, metacognitive skills are developed through activities that promote analyzing the task assignment, activating prior knowledge, goal setting, and planning at the onset of task performance (Veenman 2012). These elements are all present in the CRC method.

METHODOLOGY

This section details the CRC process, beginning with a qualitative overview of the methodology, and then, takes a close look at the purpose, implementation process, and grading scheme for each of the three phases. Note that this section presents the fully developed version of the CRC method, which incorporates all the revisions to its initial implementation done in refinement cycles; the specifics of the refinements are discussed in the later sections.

CRC Process, in Short

The CRC Crowdlearning method can be implemented easily using Google forms and group quiz software/platform such as Kahoot.

CRC is conducted in *rounds*: each completion of the three phase cycle constitutes one round. A new round of CRC is run each time that a new chunk of material is introduced to the class. After every class, or once/twice a week, students are asked to create "Who wants to be a millionaire?" type quiz questions based on the recently covered material. This type of a question – a multiple choice question – constitutes a problem statement with the answer alternatives and indication of the correct answer(s). Example questions created by the instructor (e.g., based on Week 1 material) and in-class discussions encourage questions that help students study, recall, and understand the key concepts. Students submit questions via a Google form. First, the student-created questions act as an exit ticket for the instructor to get a sense of student understanding. Second, they give the student a reason to review class notes outside of class. Students then complete a series of pairwise comparisons comprised of their peers' questions through another Google form. Lastly, the students compete in an in-class group quiz consisting of the questions that they and their peers created and that the instructor selected, perhaps after some edits. The quiz competition can be conducted offline or online, i.e., in an asynchronous or synchronous manner. The latter is best as it enables a live discussion of the material and reveals misconceptions.

CRC Process, at Length

This section provides a more in-depth explanation of the CRC stages, complete with the practically important details.

Introducing CRC to a Class Prior to the start of the first CRC round in a course, the instructor may intervene by demonstrating to the students how to create effective “Who wants to be a millionaire?” type questions. To help this cause, the first quiz competition, fully prepared by the instructor, will present *by example* the types of questions that can be posed and the level of depth/difficulty encouraged. The first ungraded, pop-up quiz competition given early in the semester engages students very well. Composed of 7–12 questions, it will take 10-15 minutes of class time accounting for very short discussions after each question (the material of the first couple of classes early in the semester tends to be light). A typical discussion includes a quick look at the fraction of the students that got the question right, an explanation of why the incorrect answers are incorrect (if they were selected by some of the students), and, if needed, a quick reminder of where/how this material was presented in the recent class(es) — this referencing activates students’ memory and creates the sense of continuity in the coverage of the material.

It is crucial that the instructor conveys to the students what constitutes a useful question, to ensure students get the most out of the CRC process. This does not have to take much class time (a lot can be explained/re-iterated in the take-home question creating assignment description), but it is effective for creating a positive rapport with the class and generating excitement about the new practice. When working with undergraduate students, more explicit instruction is advised as they may not possess the necessary skill-set to create useful and high quality questions without proper guidance. Through discussion, formal or informal, a set of guidelines should be shared about what makes for a *useful* question. The key ask is “Please try to create a question that will be useful to *you* later when you prepare for the final exam and need to remember what facts were important for this section of the material.”. It is also important to alleviate any stress by emphasizing engagement as the key goal of the CRC practice in the course and highlighting the fact that grading will largely be based on participation. Note that providing guidance for the question creation phase will also help the students to do a more thoughtful job in ranking the questions created by their peers. This guidance is imperative when incorporating this method in courses with a majority make-up of undergraduate students as they might have limited prior experience in critically evaluating their peers’ work.

Figure 1 is a flow chart of the Create-Rank-Compete process. Note that grouping students in teams for the first round of CRC can be advantageous as it promotes student-to-student engagement and allows the students to discuss the new types of assignments. In what follows, the purpose, route to implementation, and grading scheme is explained for each of the three phases of the CRC method. The *purpose* narrative states the motivation for each phase and reviews relevant literature.

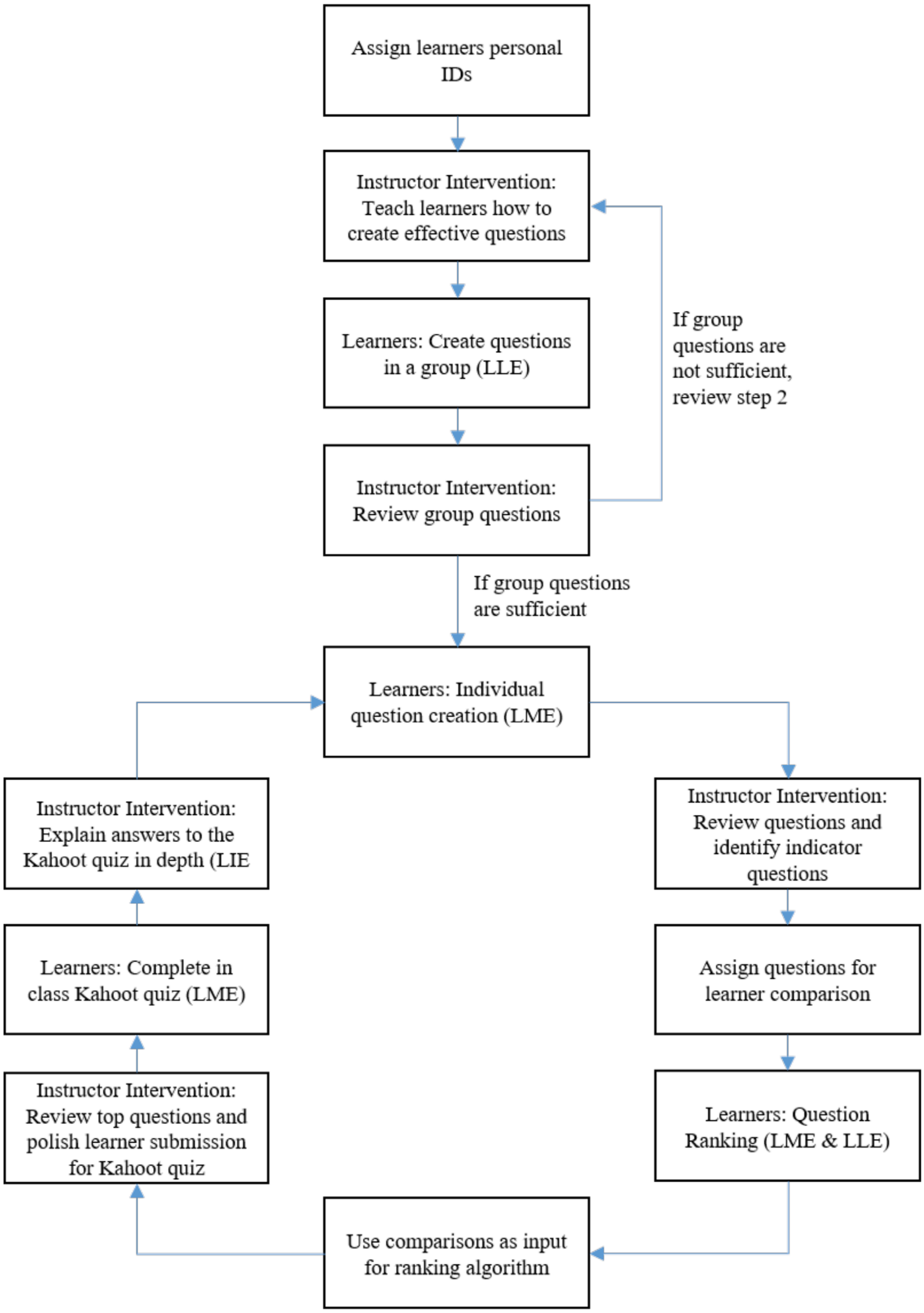


Figure 1. A flow chart of the entire Create and Rank methodology. (Note: LLE, LME, and LIE stand for learner-to-learner engagement, learner-to-material engagement, and learner-to-instructor engagement, respectively.)

The *implementation* details the technical aspects of each phase and provides examples to help others employ the method. Finally, the recommended *grading scheme* for each phase is discussed, emphasizing how it is devised to increase engagement rather than penalize students.

CRC Creation Phase: Purpose The purpose of the Creation Phase is to increase learner-to-material engagement: as students create questions, they are prompted to review course materials. Question creation is a form of alternative assessment rooted in active learning. Simply put, learning becomes active when students engage and take part in the learning process. As stated by Boud, "students are simply more likely to internalize, understand and remember material learned through active engagement in the learning process" (Boud 2013). In addition, alternative assessments have been shown to promote learner-to-material engagement and allow for a higher level of comprehension as students have a stake in their own education (Andrade-Pineda et al. 2019; Ciancimino and Cannella 2008; Sadler 1989). Furthermore, each individual student's effort facilitates and supports the progress of multiple other students.

CRC Creation Phase: Implementation Recall that in CRC, the term *question* is understood as a problem statement with the answer alternatives and indication of the correct answer(s). Questions in this form are surely known to many students (from school, SAT and GRE tests, etc.), but getting into the shoes of a *question creator* will likely be new for most of them. In any case, the use of such questions for serving the needs of a particular college course should be introduced gently.

The first question creation phase should better take place after the first class-wide quiz competition, right after the quiz or at the very beginning of the next class.

Asking the students to take initiative in creating questions, the instructor begins with a short lecture on how to think of question creation, what makes a question useful, how long and how difficult a typical question should be. Teaching the students the Bloom's taxonomy may facilitate the discussion at this stage, helping differentiate multiple questions by depth. It is also a good idea to show to the class a couple of "half-baked" questions, a couple of questions with wrong answers, and discuss the deficiencies and things to check/avoid.

Note that a question submitted with a wrong answer is not automatically a bad creation at all! For the CRC Creation phase, it is most important that students are encouraged to find and bring to the table the challenging pieces of the covered material, and express their understanding of them. It is paramount to instill confidence in students that they will not be penalized for making a mistake and, moreover, to encourage questions on the topics that need clarification. It is the questions that are too straightforward or unreasonably time-consuming that are not useful, while everything that will facilitate a well-pointed discussion and promote understanding of key course concepts is very useful. Also, it is a good idea to discourage *direct use* of bullet points found in the course slides/videos shared with the students in-class or after the class. It is good to emphasize creativity here.

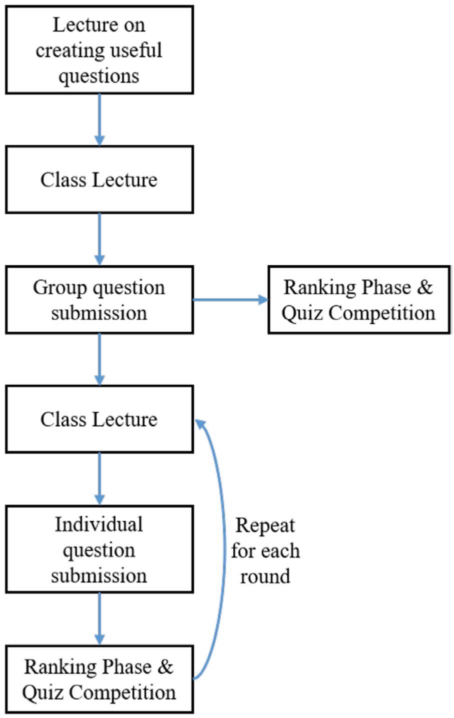


Figure 2. A flowchart depicting the steps for the Question Creation phase of the Create-Rank-Compete process.

Also, the instructor may say: "If your question can be answered by any middle-schooler who knows nothing of our subject but can read the text on the slides, then this question needs more of your intellectual work."

A flow chart displaying each step of the Creation Phase is shown in Figure 2. The Creation Phase of the CRC method is implemented as the assignment that requires students to submit 1–4 questions, via a Google Form, pertaining to the prior class' material. There are no restrictions on the type of question which can be posed, so long as it is a multiple choice question (Palmer and Devitt 2006) with 2-4 alternative answers, i.e., answer options. The creator is asked to identify which of the options they provide is the correct answer(s), however, they are permitted to forego that step and designate this question as one in which they do not know the correct answer. Allowing students to pose questions in which they themselves do not know the answer provides a greater range in questions, by depth and difficulty, while encouraging students to challenge themselves and their peers. Hereafter, the questions that were submitted by students but not yet reviewed by the instructor (in preparation for a quiz competition) are referred to as "raw" questions. It is raw questions that are used in the CRC Ranking Phase.

Example Question 1	Example Question 2
If the service rate is greater than the arrival rate, the number waiting in the queue will always be finite. a. True b. False	Which of the following is not true about Q-Q plot? a. It depends on how the data is grouped into. b. It represents quantile comparisons of the two data sets. c. Plots of the observed against expected test statistic. d. It is much better than a density- histogram when the number of data points is small.

Figure 3. Examples of student submitted (raw) questions from a course on simulation and stochastic methods.

Two examples of raw questions are displayed in Figure 3. The specific topic, which the questions pertain to, is insignificant here: our goal is to demonstrate the range in quality of questions. Example Question 1 is considered to be a high level, well polished quiz question, as it contains no errors and covers a topic at depth. Example Question 2 tests the knowledge of Q-Q plots at a deep level; it has a couple of language issues but is solid otherwise. Though not shown here, other common errors include typographical and conceptual errors, as well as incorrectly identifying the answer. Typographical errors are nonconsequential, however, conceptual errors offer a great opportunity for the instructor to refine the question and include it in the quiz. Including questions of this type allow the instructor to clear misconceptions related to the question's topic during the discussion. In addition to using the questions for the comparisons in CRC Ranking Phase and in the quizzes, the instructor may also benefit from looking over the raw creations right at submission: they can serve as an exit ticket (Danley, McCoy, and Weed 2006), i.e., inform the instructor of which points are worth emphasizing at the beginning of next class or in homework reading. Though, as class size increases, the time required to review the raw questions will increase, luckily, in large courses the instructor is typically accompanied by a teaching assistant who can aide with these tasks.

CRC Creation Phase: Grading Scheme Grades serve as an extrinsic motivation for students to study and complete assignments, however, research also shows that grades can negatively impact student well-being and intrinsic motivation (Pulfrey, Darnon, and Butera 2013; Butler and Nisan 1986; Cameron, Banko, and Pierce 2001). To ensure the students complete each question submission assignment and put effort into creating useful questions, grades are assigned based on effort and creativity, so as to encourage engagement with the learning process rather than penalize students. Our recommendation is that question creation should account for 10% of the overall course grade. Further, according to the grading scheme we adopted, which proved effective, students would stand to receive full credit for question creation if their questions fall in the top 80% of questions in at least half of the rounds; which questions are top 80% and which are bottom 20% is determined from the aggregate question ranking after the CRC ranking phase (discussed next). The recommended creation phase grading scheme's parameters (e.g., percentages used) may be adjusted depending on circumstance. The key is to incentivize students to willingly submit questions without heavily penalizing them for variable participation from round to round. The students should know that they can skip a round if desired; in fact, experience showed the participation levels are much higher than 50% in all phases (see Table 1). Indeed, full CRC participation achieves very high engagement with the course material, so the instructor should be mindful of the total workload the student has overall during the semester. The importance of the latter consideration became evident during the Zoom-heavy, Covid-affected school semesters, when students felt overloaded and found it difficult to concentrate and motivate themselves to complete work for online courses (Nambiar 2020).

CRC Ranking Phase: Purpose The purpose of this phase is to establish a new form of peer engagement, as well as facilitate learner-to-material engagement. As a homework assignment, each student is asked to analyze 20 unique peer-created questions, presented in pairs, facilitating access to the covered subject matter concepts. Questions are compared pairwise as this is the simplest, natural form of soliciting human judgement (Chen et al. 2013). In this phase, students are engaging with each other in a blind manner, which in part informs them of each other's progress. Indeed, Cannella et al. find that students reach a higher level of learning when being questioned by peers than when the class is instructor-focused, so by reviewing peer questions students obtain a greater depth of understanding of the material (Cannella, Ciancimino, and Campos 2010).

CRC Ranking Phase: Implementation The question ranking phase is the most complex of the three phases, consisting of three stages: (1) comparisons are assigned to the students, (2) students complete their assigned comparisons, and (3) the comparisons are fed to a ranking algorithm which determines the rank weight and position for each question — the more voted-for questions assume higher positions on the list. See Figure 4.

In stage 1, questions are assigned to students so as to enhance their exposure to material: each question (of the 20) must be unique, maximizing the amount of information gathered, and also, informing the ranking algorithm. The exact number of comparisons to assign to each student is set so as to ensure that each question is reviewed multiple times. Note that the comparisons are blind: it remains unknown to the students who ranks whose question(s).

To ensure that students conduct the comparisons honestly, three pairs of *indicator questions* are assigned. Each such "ground truth" comparison should be composed of one very-high-utility (i.e., "best") question and one very-low-utility ("worst") question. To this end, after the students submit questions, the instructor reviews them and chooses the indicator questions. Alternatively, the instructor may contribute some or all of their own specially created questions toward this purpose. Any one of the "best" questions, compared to any one of the "worst" questions, should be easy to

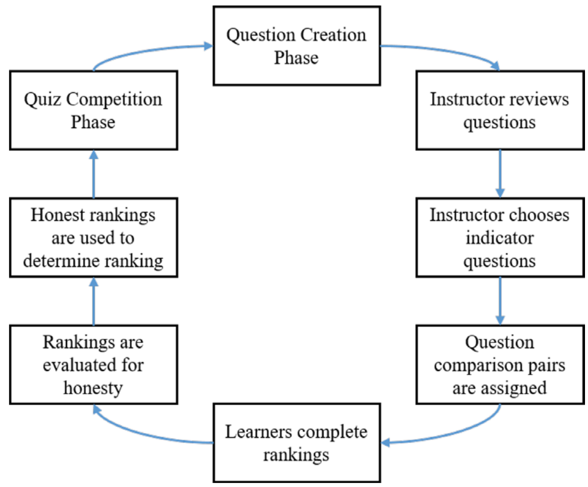
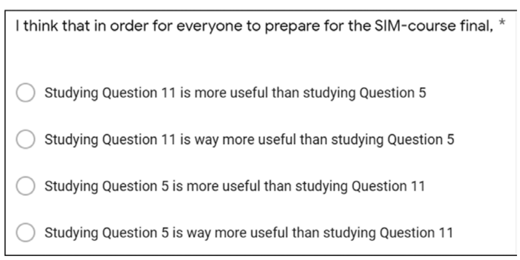


Figure 4. A flowchart depicting the steps for the CRC Question Ranking phase.

recognize as the question more useful for studying. To allot for natural error in judgment, a student needs to correctly rank two of the three indicator pairings. If a student ranks correctly at least $\frac{2}{3}$ of the ground truth comparisons, then their entire rankings are considered to be honest. Only honest rankings are subsequently used as input to the rank aggregation algorithm.

In stage 2, students complete their assigned comparisons. Comparisons can be easily administered via a Google Form, formatted as displayed in Figure 5. In this format, question quality is assessed only relatively, determined by perceived usefulness: it allows students to express the level of confidence in their choice by claiming that one of the questions is "... way more useful to study..." versus just "... more useful to study...".



I think that in order for everyone to prepare for the SIM-course final, *

- Studying Question 11 is more useful than studying Question 5
- Studying Question 11 is way more useful than studying Question 5
- Studying Question 5 is more useful than studying Question 11
- Studying Question 5 is way more useful than studying Question 11

Figure 5. Example of a comparison assignment given to students in the CRC question ranking phase.

In stage 3, the data collected from the question comparisons are used as an input for the ranking algorithm. The algorithm uses a Markov Chain logic-based ranking method to calculate the rank weights for each question. The weights determine the rank position, i.e., the order of the questions from most useful to least useful as perceived collectively by the students. This order is used by the instructor when creating the in-class Kahoot quizzes. This means that the student whose question has been voted in as useful by many peers will be more likely to see their question be used in the quiz competition.

We acknowledge the fact that not all adopters of CRC will have access to algorithms that facilitate comparison assignment and rank aggregation. Not to worry, questions may be assigned manually, in Microsoft Excel, or at random. As for the aggregation, the ranking can be done via a point system, that is, a question receives (+1) or (+2) points each time it is favored more or way more, respectively, and when a question is unfavorable it receives (-1) or (-2) points, depending on intensity. Once points have been allotted, the questions can be ordered from greatest number of points to least number of points, in any software that enables working with spreadsheets, such as Excel. However, if an adopter is interested in using the comparison assignment model or ranking algorithm, they are discussed at length in the Appendix section. Assigning comparisons manually to a large class of students is cumbersome, thus, if the CRC method is being adopted in a large course we recommend utilizing the provided algorithms for implementing the question ranking phase. The algorithms automate the assignment of questions and the ranking of questions, allowing this method to be scalable in large courses at both the graduate and undergraduate level.

CRC Ranking Phase: Grading Scheme The question ranking phase accounts for 10% of the students' overall course grade. Students who contribute honest rankings, i.e., rankings in which at least $\frac{2}{3}$ of the indicator pairings are correctly ranked at least 50% of the time (in half of all the CRC rounds), receive full credit for this phase.

CRC Quiz Competition Phase: Purpose The purpose of the CRC competition phase is to increase learner-to-material and learner-to-instructor engagement via formative assessment, gamification, and situation-driven lecturing. Composed of polished versions of student-submitted raw questions, a class-wide Kahoot quiz provides students with immediate feedback in the way of formative assessment. This phase exploits a from of gamification, which is known to make learning more effective (Sailer and Homner 2020). Additionally, the instructor provides in-depth explanations of each question with the help of the students, which is a form of active learning that facilitates learner-to-instructor engagement. This also provides the instructor with a structured manner in which to review key concepts and bridge knowledge gaps. Indeed, many instructors like to refresh students' memory at the beginning of every class to put them "on track" to digest more material. CRC quiz competitions serve this purpose well.

CRC Quiz Competition Phase: Implementation Once the question rankings are established, the instructor reviews the entire set of the questions submitted in the round. Note that the Kahoot online platform can take a spreadsheet with questions directly as an input for a quiz, thus we recommend having students submit questions via a Google form that has the same structure that Kahoot requires. This allows an instructor to automate the quiz generation, which is especially useful in large courses. Editing questions in the spreadsheet may require simplistic changes such as fixing grammatical errors, or more complex changes in which the question remains the same conceptually but the language is heavily altered. The instructor also reviews the provided answers to guarantee the correct answer is present and designated as such. Per the instructor's judgement, a set of questions is chosen for a live (in-class in a physical classroom or online) quiz competition or an offline quiz competition which may or may not require synchronous participation.

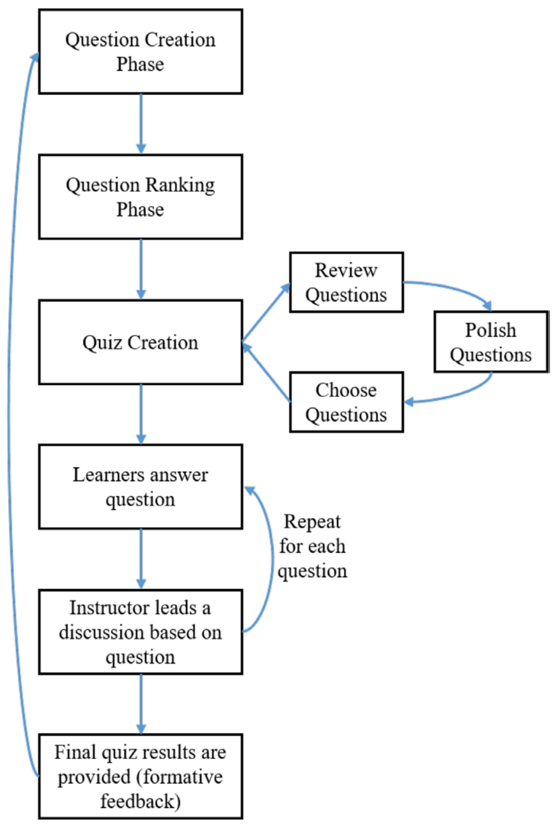


Figure 6. A flowchart depicting the steps for the Quiz Competition phase of the Create-Rank-Compete process.

It is important to emphasize that not only the "best" raw questions should be considered for the quiz, but also, questions which expose students' misunderstanding. Referring back to Figure 3, Example Question 1 was excellent at submission, while Example Question 2 was much underdeveloped, however, both are equally useful in a quiz competition. Once polished, Question 2 will allow the professor to fill many knowledge gaps that students apparently have. Note that while preparing a quiz, it is a good idea to take note of all the raw questions that touch on similar or the same particular knowledge points. The instructor may find that either multiple such questions can be folded into one or two, or that a certain definition or concept is misused in many questions. This misuse would signal the need to introduce a new question that will highlight the particular definition.

A quiz competition takes anywhere between 15 minutes and the full length of the class. The recommended length is 25 minutes, which includes the time for discussion after each question. In quiz competitions that we conducted in Kahoot, we typically gave students either 30 or 60 seconds to answer a question, and very rarely 90 seconds.

It is convenient to use Kahoot as a platform for quiz competitions, because all the questions can be fed into Kahoot directly as a spreadsheet. Since Google Forms can aggregate all questions into a single spreadsheet for download, then the ordering, polishing of raw questions, and any addition/removal of questions can be done in the same spreadsheet. Then, the spreadsheet can be instantaneously uploaded onto Kahoot, after a very fast and straightforward formatting, — and the quiz is ready to run for the entire class. Note that once completed, a Kahoot quiz competition may be given to students for another (offline) run — immediately after the original (online) run, or a bit later, or at the end of the semester to help students with preparation for exams.

CRC Quiz Competition Phase: Grading Scheme In a quiz competition (e.g., run in Kahoot), students compete against each other to score the most points. This gamified quiz serves as a formative assessment for students to gauge their knowledge and see how they compare to their peers. However, in order to motivate students to study for quizzes, average performance on the quizzes is used as a summative assessment. Kahoot quizzes account for 10% of the overall course grade. Students receive full credit for the Kahoot quiz portion of their grade if they rank in the top 80% on at least half of the quizzes (CRC rounds).

In our initial CRC implementation, the average point separation between students in the Kahoot quizzes was 491.25 points, which is less than half the points allotted per quiz question (1000). This demonstrates that students are easily able to move up in the ranks, meaning that if they put effort into the quizzes, then it will not be difficult to rank in the top 80% on at least half the quizzes. This ensures the grading schema is fair and rewards students based on effort. Similarly to the grading scheme for the question creation component, this is used to motivate students but not to penalize them, thus emphasizing engagement with the learning process rather than correctness.

RESULTS AND TAKEAWAYS

The CRC method has been implemented in a series of courses since the Spring semester of 2020. For the purposes of this manuscript we will focus on the initial implementation in a course on simulation and stochastic modeling, which will be referred to as Sim-Course throughout this section. The initial implementation of CRC illuminated areas in which the method was achieving the established goals and places which required improvement. Outlined within this section are key pieces of feedback obtained from students which guided changes in the process, a discussion on the impact the method had on student grades, and methods for evaluating student honesty during the ranking phase of CRC.

Student Feedback and Participation

During the initial CRC implementation, students completed a survey created by the researchers at the close of the first implementation. Note that this survey was not used as an instrument, but rather used to gain insights into students' perception of the method to determine if future implementations would be considered. Provided within this section is student feedback which informed future iterations of the CRC method. The following student quotes are from student responses to two questions, (1) "What was the most beneficial part of this process for you?" and (2) "If you could change anything about this process, what would you change and why?".

Quotes 1-4 were provided in response to the question, "What was the most beneficial part of this process for you", referring to the CRC process. The key takeaway from the first quote is in bold, the student felt comfortable vocalizing their doubts and questions. This notion is supported by the literature, that tasking students with question creation facilitates an inquisitive classroom environment in which students are more apt to ask questions when they do not understand and promotes curiosity (Chin 2002).

> Student 1, Compared to other courses that i have taken, i felt this class was really helping. It was always very easy to connect with the Professor and the TA, even at a short notice. Most importantly, **Professor made us feel very comfortable in asking doubts** which were cleared in an understanding way. I will really miss this method of teaching. The highlighting part of this course was the crowdlearning method. I would like to have such experiences in my educational career.

The second, third, and fourth student quotes point to the structure of the CRC process, which is a compressed Bruners' spiral curriculum. In essence, a spiral curriculum revisits the same topics with increased levels of complexity which allow students to incrementally build on their prior knowledge. Student 2, "It helps me to revising the basic concept of the course over and over again." Student 3, "Quizzes and question creations helped me keep in touch with the course through out the semester." Student 4, "Creating our questions helped us keep in touch with the subject concepts throughout the semester." These quotes demonstrate that the structure of the method is achieving the goal of reinforcing the key concepts and course topics.

Additionally, students were asked, "If you could change anything about this process, what would you change and why?" The following responses were considered when revising the method for future iterations. Student 5, "Question preparation needs to be reduced to one time a week. Otherwise this will be lot of work load and would affect quality of questions." Student 5's sentiment is reiterated by students 6 and 7, as well as multiple others. Student 6, "The question creation could be made once a week." Student 7, "Question creation frequency should be less." The comments provided by students 5, 6, and 7 prompted us to decrease the frequency of CRC rounds from two rounds a week to one round per week.

The last type of comment referred to the concept of students conducting the question ranking. Student 8, "Questions should be ranked by professor and TA to avoid partiality amongst student grading." Note that this comment was made by two other students. This signals that the students do not understand the purpose of comparing their peers' created questions, which prompted a change to the way CRC is introduced to students, as described in the previous section. In future iterations, the CRC process is reviewed and the purpose for each phase is explained to the students. The purpose of the ranking phase is to expose students to more content and to provide them opportunities

Table 1. Participation data from the Spring 2020 implementation of the Create-Rank-Compete method in a course on simulation and stochastic processes. Note that the class had a total of 31 students.

Student Participation in Sim-Course (Spring 2020) (n = 31)			
	Question Creation	**Question Ranking**	**Kahoot Quiz**
Round 1	31	31	29
Round 2	31	31	31
Round 3	31	31	31
Round 4	31	31	29
Round 5	31	29	29
Round 6	28	30	31
Round 7	31	29	31
Round 8	30	30	29
Round 9	30	30	27
Round 10	30	28	27
Round 11	31	31	21
Round 12	31	31	26
Round 13	28	28	28
Round 14	27	25	27

to actively engage in the course outside of class time, this is now clearly explained to students to avoid misunderstandings.

Impact on Student Grades

Throughout the Sim-Course, students must complete three projects and a final exam, in addition to the CRC requirements. During the 2019 and 2020 offerings of the course, the same final examination was used and the only change to the curriculum was the introduction of the CRC method. Thus, one way to establish the credibility of the CRC method is to compare final exam and course grades between the 2019 and 2020 offerings. Figure 7 displays distributions of student final exam and course grades, as well as the distributions of their cumulative GPAs over their course of the study at the university (to reveal biases, if any).

The left-hand-side graph in Figure 7 displays the final exam and final course grade distributions for the Sim-Course in 2019 and 2020. In 2020, the final exam grades are significantly greater than the grades from 2019, as shown in Table 2. It is important to note that the same exam was used in both semesters, though the order of the questions differed, and the students in each offering have similar academic standing, as shown in the graph on the right-hand-side of Figure 7. We tested the claim that the cumulative GPAs of students in 2019 and 2020 were not equal; there was not sufficient evidence to support this claim (as shown in Table 2). Thus, the increase in final exam grades between 2019 and 2020 cannot be attributed to an easier examination nor to the students being of a higher academic caliber. We can deduce that the increase in final exam grades is most likely

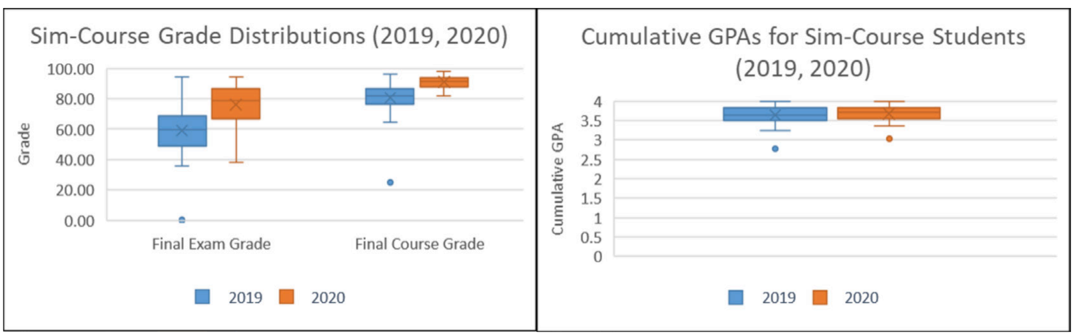


Figure 7. The graph on the left-hand-side shows the grade distributions for the final exam and for the course as a whole for students in the Sim-Course across 2019 and 2020. The graph on the right-hand-side shows the cumulative GPA distributions for the same set of students over their entire program of study at the university.

Table 2. This table displays the results from three Welch's t-tests with unequal variances in which the following samples were compared across years 2019 and 2020: Sim-Course final exam grades, Sim-Course final course grades, and cumulative GPAs. Statistically significant values are in bold and indicated by a (*).

	Sim-Course Final Exam Grades		Sim-Course Final Course Grades		Sim-Course Cumulative GPAs	
Year	2019	2020	2019	2020	2019	2020
Mean	59.17	76.22	80.74	91.07	3.65	3.68
Variance	268.28	159.43	117.80	17.33	0.06	0.04
Observation	44.00	32.00	44.00	32.00	44.00	32.00
Hypothesized Mean Difference	0.00		0.00		0.00	
df	74.00		59.00		72.00	
t Stat	–5.12		–5.76		–0.64	
P (T<=t) one-tail	0.0000*		0.0000*		0.26	
t Critical one-tail	1.67		1.67		1.67	
P (T<=t) two-tail	0.0000*		0.0000*		0.53	
t Critical two-tail	1.99		2.00		1.99	

due to the incorporation of the CRC method as it is the only substantial change between the two course offerings in these subsequent years.

Note finally, that the duration of the Master's program in the studied department is 1.5 years, with each new cycle beginning in Fall, so communication with prior year's students about the final exam is all but impossible between the Spring student populations across two consecutive years.

Similarly, the final course grades for students in the 2020 semester are significantly greater than the grades of students in 2019. The assignments and grader in the course were consistent and the course was made up of the similar student populations. Note that some of the grade increase may be due to a grade bump resulting from the CRC method as students were predominantly graded based on effort and all did well in CRC grade-wise. However, the most difficult course components - projects and exam - accounted for the highest portion of the grade (compared to the labs and CRC components) in both years. Besides, even if the CRC grade portion may have had any effect on the absolute grade values, this cannot be the case with the exam grades.

Thus, we conclude that the CRC method helped improve student learning and understanding of course material, which in turn positively affected their course grades.

Assessing and Monitoring Student Effort

The CRC method is predicated on iterative interaction with student-generated questions, as such it is vital that these interactions be monitored and assessed to gauge the authenticity of responses and measure student effort. In order to increase learner-to-material engagement and improve the quality and accuracy of the rankings, we implemented two methods, (1) assessing learner responsibility via question quality and quiz scores and (2) incorporating indicator question to weed out dishonest comparisons. Assessing learner responsibility allows us to analyze already completed rankings with dishonest comparisons whereas incorporating indicator questions decreases the prevalence of dishonest comparisons. Note that in the final CRC method discussed in the "Methodology" section and displayed in Figure 1, the use of indicator questions is favored over the method for assessing learner responsibility. However, in cases in which indicator questions were not used, the method described for assessing learner responsibility allows the instructor to post-process comparisons. To demonstrate the validity of implementing indicator questions two metrics were used, percentage point differentials in honest responses and a measure of inter-rater reliability.

Assessing Learner Responsibility

The process of assessing learner responsibility is performed post-completion of the CRC method on comparisons which appear to be dishonest. This is vital for ensuring accuracy within the question rankings as well as to grade students based on engagement with the learning process rather than strictly on completion. Data collected during phases I and III of the CRC process is used when assessing learner responsibility, namely the two factors determining a students' responsibility score are question creation scores and Kahoot quiz scores.

The first factor in a students' responsibility score is their question creation scores. We devised a method for comparing raw and polished questions in which students earn points for creating questions that are incorporated into the Kahoot quizzes. Each question was reviewed by two content-area specialists and give a score based on the quality of the question. Raw questions which require minor revisions in order to be accepted receive 1 point and those which require major revisions earn 0.5 points. Examples of such questions is exhibited in Figure 8. A students' question creation score is the summation of their scores in every round, assuming learner responsibility is assessed after the final round of CRC has been administered. This enables us to analyze student progress throughout each round and in the semester as a whole.

The second factor comprising a students' responsibility score is devised from Kahoot quiz scores. Similarly to the way in which we encoded the raw and polished questions, we encoded student Kahoot quiz scores. For each round, the students who finished in the top five on the Kahoot quiz received 1 point and the students in positions six through ten earned 0.5 points. These scores were summed across all rounds to determine a student's total Kahoot score; this score made up made up half of their responsibility score.

Example Raw and Polished Question Pairings
for Varying Encoded Scores (0, 0.5, and 1)

Example Raw Question with Score 1 What two values are required to form the confidence interval? a. 5^{th} and 95^{th} percentile b. simulated mean and st. deviation c. Error level and calculated number of iterations Correct Answer: b	Example Polished Question with Score 1 What two things are required to build a confidence interval? a. The 5^{th} and 95^{th} percentile b. The 50^{th} percentile and standard deviation c. Level of confidence and number of iterations d. Practical significance and statistical significance Correct Answer: c
Example Raw Question with Score 0.5 If, after a large number of replications, the CI for difference in outputs for system design A vs. B is 30 +/- 40, then a. A and B are statistically indistinguishable b. A and B are practically indistinguishable c. If more is better, then A is better than B d. We cannot state any of the above Correct Answer: c	Example Polished Question with Score 0.5 If after almost INFINITE replications, a difference in measure of model a. The result is statistically insignificant b. The models are virtually indistinguishable w.r.t this measure c. Then we assume that the model with the better mean metric is better Correct Answer: b
Example Raw Question with Score 0 When is simulation not advised to use – a. To understand different capabilities of a machine b. When we do not have enough realistic data c. To verify analytical solution d. When new policies are needed to be implemented Correct Answer: c	

Figure 8. (Row 1) A pair of example raw and polished student-generated questions with the coded score of 1, meaning the question was accepted with minor revisions. (Row 2) A pair of example raw and polished student created questions with the coded score of 0.5, meaning the question was accepted with major revisions. (Row 3) An example of a raw question with coded score 0, meaning the question was rejected.

Students with high scores from the question submission process are capable of creating beneficial questions and are therefore more apt to determine which questions are useful. In the same vein, students who consistently scored well on the Kahoot quizzes display a mastery of the material and are more capable of interpreting which questions are most useful. Thus, these two factors are integral in assessing learner responsibility.

A learner's responsibility score equates to the sum of the scores from the question creation phase and the Kahoot quiz competitions. Once each student's responsibility score is determined, the third quartile is calculated and used as the lower threshold or cut-off point. All students with scores greater than the threshold are placed in the responsible learner group. When utilized prior to the end of the course, this enables a professor to identify students who consistently fall below the cutoff point, i.e. students who may need instructor intervention. This benefits both the instructor and the student as the instructor is informed of where to expend energy and the students are provided with additional help and guidance. Once the responsible learners are identified, the rankings from all rounds are rerun using only the comparisons provided by the responsible learner group.

When we employed this method using the data collected from the initial implementation the results, though more promising than the rankings determined via the entire class, were still wanting. This led us to devise a method to signal untrustworthy comparisons prior to completion of the method and to incentivize students to give honest comparisons.

Implementing Indicator Questions

In order to increase learner-to-material engagement and improve the quality and accuracy of the rankings, we incorporated indicator question to decrease the prevalence of dishonest comparisons.

Upon analyzing the results from students question rankings during the initial implementation, it became evident that some students did not complete the comparisons honestly. Thus, we developed a method that would ensure trustworthiness of the question comparison outcomes. Indicator questions and a grading scheme in which students are graded on effort and the honesty of their comparisons were incorporated into the method.

Following the success from the initial implementation, the CRC method was adapted for the Fall 2020 semester in a course on Social Network Behavior Analysis, referred to as SNA-Course, which was taught by the same professor as the SIM-course. As previously discussed, the method needed to adopt a way of determining trustworthiness within the comparisons and incentivizing students to conduct caparisons honestly. Thus, we incorporated indicator questions and a grading scheme in which students are graded on effort and the honesty of their comparisons. Note that indicator questions are discussed in the "Methodology" section, though they were not a part of the initial design, but were added following the first implementation of the CRC method.

The indicator questions used are determined via an expert, i.e. the instructor. Five of the best and five of worst questions, as determined by the instructor, are used as indicator questions to ascertain which comparisons are truthful and which are not. Each student is randomly assigned three pairs of questions comprising of one 'best' question and one 'worst' question.

These comparisons are devised in such a way that if a student were to complete the comparisons honestly, then they should choose the better question at least 2 out of 3 times, leaving room for honest student error. If a student chooses the worst question a majority of the time, then their comparisons are marked as untrustworthy and removed to improve the accuracy of the rankings and to inform grading.

Indicator questions were first introduced in week 5 of the Fall semester and 47.62% of students correctly ranked at least 2 out of 3 of the indicator questions signalling that a majority of the students were not performing comparisons honestly. Upon this realization, the instructor reminded the students that they are graded based upon how many rounds of honest comparisons they complete and informed the students of the presence of indicator questions. When analyzing the week 7 comparisons, we found that 60% of students correctly ranked at least 2 out of 3 of the indicator questions. Thus, by including indicators questions and using grades as incentivization we were able to increase the percentage of honest comparisons by approximately 12.38 percentage points, which in turn increases learner engagement with material and the accuracy of the question rankings. As the rankings inform the instructor on which questions to include on quizzes, increasing the accuracy of the rankings improves the material on the quizzes and the in class review sessions benefiting the overall teaching and learning processes.

The second metric for establishing the validity of implementing indicator questions is inter-rater reliability (IRR). IRR represents the amount of agreement between raters, accounting for agreement that happens by chance. This is important as there is no ground truth from which to evaluate student rankings, thus IRR provides a metric for determining the quality of rankings. For our purposes, each student's responses were compared to the aggregate ranking to determine the level of agreement between each rater and the class as a whole. We found that the average IRR value, also referred to as a Kappa value, for week 5 is 0.57 and the average IRR value for week 7 is 0.67; this equates to a difference of 0.1, but what does this mean? Figure 9 displays the standards for determining the quality of IRR from four widely-cited articles (Landis and Koch 1977; Cicchetti and Sparrow 1981; Fleiss et al. 1981; Regier et al. 2013).

As there is no set standard for comparing inter-rater reliability values, we will use a variety of standards as set by Landis & Koch, Cicchetti & Sparrow, Fleiss, and Regier et al, displayed in Figure 9. The Kappa value for week 5 is 0.57, this value falls in yellow range across the board, indicating a fair/good value. Whereas the Kappa value for week 7 is 0.67, which lands in the blue range indicated a good or very good value, with the exception of the standard set by Fleiss in which it would be considered "Fair to Good". This demonstrates that implementing indicator questions (and informing students) increases the inter-rater reliability value substantially.

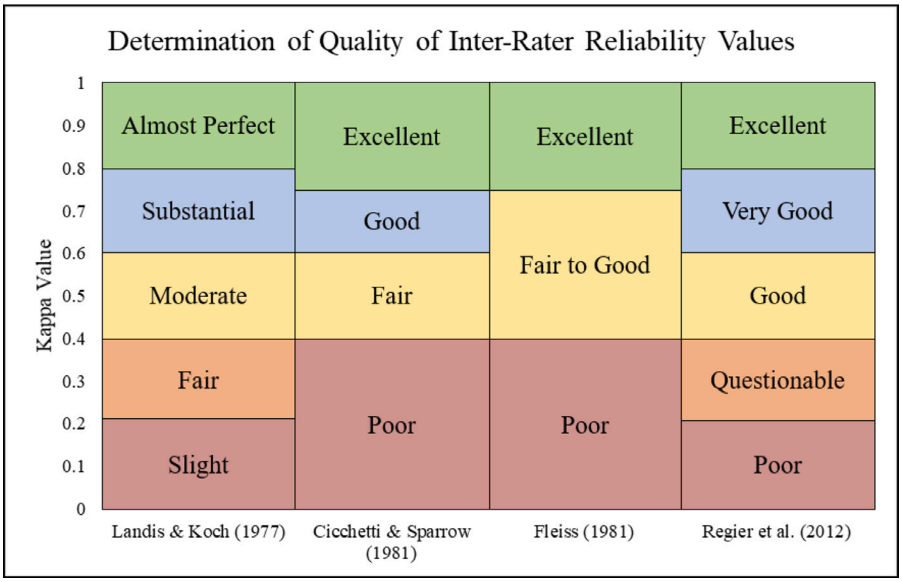


Figure 9. An adaptation of an image created by Eric Youngstrom depicting the quality of inter-rater reliability values (Kappa values) as established in four widely cited articles (Landis and Koch 1977; Cicchetti and Sparrow 1981; Fleiss et al. 1981; Regier et al. 2013). Note that Kappa values range from 0 to 1.0.

CONCLUSION

This paper outlines a new method of teaching, Create-Rank-Compete, which can be implemented in in-person classes or distance learning models. This method serves as a just-in-time solution for globalized virtual learning, the demand of which has heavily increased due to the Covid-19 pandemic. Through student participation data, qualitative feedback provided in surveys, and students' graded we have demonstrated that the CRC method is accepted by students, incentivizes them to engage with course material, and has a positive impact on not only their exam grades, but their overall course grades. Further, the CRC method is beneficial to the teaching-learning process as it forms a three step cycle: (1) students create questions, (2) students interact with their peers creations and information gleaned from this step informs the instructor on the state of students' knowledge, and (3) there is a structured review of course material based on the information gathered by the instructor in step two.

Create-Rank-Compete is the latest evolution of our Crowdlearning method, which has now been used for six years in an engineering department at a large public university in the northeastern United States. The CRC method was piloted in small, cross-listed courses predominantly consisting of graduate students. We recognize that this is a limitation of our study as implementation in large courses may pose additional challenges, such as requiring a greater time commitment from the instructor or teaching assistants that manage the Crowdlearning. To mitigate this, we recommend using a Google form to collect question submissions and organizing the form to be consistent with Kahoot to streamline the quiz creation process. Further, we have included two algorithms with explanations for implementation in the Appendix section. These will reduce the time required to assign pairwise comparisons to students, arguably the most time consuming step in the process, and automates the rank aggregation of questions. Lastly, we urge adopters who use this method with undergraduate students spend more time teaching students how to create high quality questions and providing ample examples.

This method will continue to be used in the engineering department with the goal of expanding its adoption. More specifically, we plan to use the CRC method in a diverse set of engineering courses, ranging from small, graduate courses to large undergraduate courses. This will allow us to continue with the refinement of this method. Further, we will pilot additional activation phases with the purpose of identifying which activity phase is preferred by students and which garners the best results. In particular, we will evaluate the incorporation of a question editing activation phase, in which students may be tasked with editing their peers' creations collaboratively and individually.

REFERENCES

Aalst, Jan van. 2009. "Distinguishing Knowledge-Sharing, Knowledge-Construction, and Knowledge-Creation Discourses." *International Journal of Computer-Supported Collaborative Learning* 4 (3): 259–87.

Andrade-Pineda, J L, J M Leon-Blanco, M Calle, and P L Gonzalez-R. 2019. "Project Management at University: Towards an Evaluation Process Around Cooperative Learning." *World Academy of Science, Engineering and Technology International Journal of Educational and Pedagogical Sciences* 13 (1): 32–36. https://doi.org/0000000091950263.

Beal, Carole R, Paul R Cohen, et al. 2012. "Teach Ourselves: Technology to Support Problem Posing in the STEM Classroom." *Creative Education* 3 (04): 513.

Boud, David. 2013. *Enhancing Learning Through Self-Assessment*. Routledge.

Braha, Dan, and Oded Maimon. 1997. "The Design Process: Properties, Paradigms, and Structure." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 27 (2): 146–66.

Bransford, John, John D Bransford, Ann L Brown, and Rodney R Cocking. 1999. *How People Learn: Brain, Mind, Experience, and School*. National Academies Press.

Brindley, Clare, and Susan Scoffield. 1998. "Peer Assessment in Undergraduate Programmes." *Teaching in Higher Education* 3 (1): 79–90.

Butler, Ruth, and Mordecai Nisan. 1986. “Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance.” *Journal of Educational Psychology* 78 (3): 210.

Cameron, Judy, Katherine M Banko, and W David Pierce. 2001. “Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues.” *The Behavior Analyst* 24 (1): 1–44.

Cannella, Salvatore, Elena Ciancimino, and Mónica López Campos. 2010. “Mixed e-Assessment: An Application of the Studentgenerated Question Technique.” In *IEEE EDUCON 2010 Conference*, 769–73. IEEE.

Chen, Xi, Paul N Bennett, Kevyn Collins-Thompson, and Eric Horvitz. 2013. “Pairwise Ranking Aggregation in a Crowdsourced Setting.” In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, 193–202.

Chin, Christine. 2002. “Student-Generated Questions: Encouraging Inquisitive Minds in Learning Science.” *Teaching and Learning* 23 (1): 59–67.

Ciancimino, Elena, and Salvatore Cannella. 2008. “The Alternative Assessment Experiment.” *International Journal of Learning* 15 (2).

Cicchetti, Domenic V, and Sara A Sparrow. 1981. “Developing Criteria for Establishing Interrater Reliability of Specific Items: Applications to Assessment of Adaptive Behavior.” *American Journal of Mental Deficiency*.

Danley, Angela, Ann McCoy, and Rahila Weed. 2006. “Exit Tickets Open the Door to University Learning.” *InSight: A Journal of Scholarly Teaching* 11: 48–58.

Disessa, Andrea A. 2002. “Why ‘Conceptual Ecology’ Is a Good Idea.” In *Reconsidering Conceptual Change: Issues in Theory and Practice*, 28–60. Springer.

Elliot, Andrew J. 2020. “Competition and Achievement Outcomes: A Hierarchical Motivational Analysis.” *Motivation Science* 6 (1): 3.

English, Lyn D. 1998. “Children’s Problem Posing Within Formal and Informal Contexts.” *Journal for Research in Mathematics Education* 29 (1): 83–106.

Farasat, Alireza, Alexander Nikolaev, Suzanne Miller, and Rahul Gopalsamy. 2017. “Crowdlearning: Towards Collaborative Problem-Posing at Scale.” In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, 221–24.

Fleiss, Joseph L, Bruce Levin, Myunghee Cho Paik, et al. 1981. “The Measurement of Interrater Agreement.” *Statistical Methods for Rates and Proportions* 2 (212-236): 22–23.

Fosnot, Catherine Twomey, and Randall Stewart Perry. 1996. “Constructivism: A Psychological Theory of Learning.” *Constructivism: Theory, Perspectives, and Practice* 2 (1): 8–33.

Garcia, Stephen M, Avishalom Tor, and Tyrone M Schiff. 2013. “The Psychology of Competition: A Social Comparison Perspective.” *Perspectives on Psychological Science* 8 (6): 634–50.

Gibbs, Brian C. 2014. “Reconfiguring Bruner: Compressing the Spiral Curriculum.” *Phi Delta Kappan* 95 (7): 41–44.

Gunstone, Richard F., and Ian J. Mitchell. 2005. “Chapter 5 - Metacognition and Conceptual Change.” In *Teaching Science for Understanding*, edited by Joel J. Mintzes, James H. Wandersee, and Joseph D. Novak, 133–63. Educational Psychology. Burlington: Academic Press. https://doi.org/https://doi.org/10.1016/B978-012498360-1/50006-4.

Johnston, Howard. 2012. “The Spiral Curriculum. Research into Practice.” *Education Partnerships, Inc.*

Landis, J. Richard, and Gary G. Koch. 1977. “An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement Among Multiple Observers.” *Biometrics* 33 (2): 363–74. http://www.jstor.org/stable/2529786.

Li, Ming-Chaun, and Chin-Chung Tsai. 2013. “Game-Based Learning in Science Education: A Review of Relevant Research.” *Journal of Science Education and Technology* 22 (6): 877–98.

Linn, Marcia C, Douglas Clark, and James D Slotta. 2003. “WISE Design for Knowledge Integration.” *Science Education* 87 (4): 517–38.

Masters, Kenneth, and Trevor Gibbs. 2007. "The Spiral Curriculum: Implications for Online Learning." *BMC Medical Education* 7 (1): 1–10.

Mishra, Shitanshu, and Sridhar Iyer. 2015. "An Exploration of Problem Posing-Based Activities as an Assessment Tool and as an Instructional Strategy." *Research and Practice in Technology Enhanced Learning* 10 (1): 1–19.

Morris, Besty. 2020. "Why Does Zoom Exhaust You? Science Has an Answer." *Wall Street Journal, March* 27.

Musingafi, Maxwell Constantine Chando, Kwaedza Enety Muranda, et al. 2014. "Students and Questioning: A Review of the Role Played by Students Generated Questions in the Teaching and Learning Process." *Studies in Social Sciences and Humanities* 1 (3): 101–7.

Nambiar, Deepika. 2020. "The Impact of Online Learning During COVID-19: Students' and Teachers' Perspective." *The International Journal of Indian Psychology* 8 (2): 783–93.

Negahban, Sahand, Sewoong Oh, and Devavrat Shah. 2017. "Rank centrality: Ranking from pairwise comparisons." *Operations Research* 65 (1): 266–87. https://doi.org/10.1287/opre.2016.1534.

Palmer, Edward, and Peter Devitt. 2006. "Constructing Multiple Choice Questions as a Method for Learning." *Annals-Academy of Medicine Singapore* 35 (9): 604.

Palmero, Julio Ruiz, and José Sánchez Rodríguez. 2012. "Peer Assessment in Higher Education. A Case Study." *Czesław Kupisiewicz*, 247.

Pulfrey, Caroline, Céline Darnon, and Fabrizio Butera. 2013. "Autonomy and Task Performance: Explaining the Impact of Grades on Intrinsic Motivation." *Journal of Educational Psychology* 105 (1): 39.

Purchase, Helen C. 2000. "Learning about Interface Design Through Peer Assessment." *Assessment & Evaluation in Higher Education* 25 (4): 341–52.

Regier, Darrel A, William E Narrow, Diana E Clarke, Helena C Kraemer, S Janet Kuramoto, Emily A Kuhl, and David J Kupfer. 2013. "DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses." *American Journal of Psychiatry* 170 (1): 59–70.

Rotherham, Andrew J, and Daniel Willingham. 2009. "21st Century." *Educational Leadership* 67 (1): 16–21.

Sadler, D Royce. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–44.

Sailer, Michael, and Lisa Homner. 2020. "The Gamification of Learning: A Meta-Analysis." Springer.

Shute, Valerie J, and Robert Torres. 2012. "Where Streams Converge: Using Evidence-Centered Design to Assess Quest to Learn." *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research* 91124.

Silver, Edward A, Joanna Mamona-Downs, Shukkwan S Leung, and Patricia Ann Kenney. 1996. "Posing Mathematical Problems: An Exploratory Study." *Journal for Research in Mathematics Education* 27 (3): 293–309.

Topping, Keith. 1998. "Peer Assessment Between Students in Colleges and Universities." *Review of Educational Research* 68 (3): 249–76.

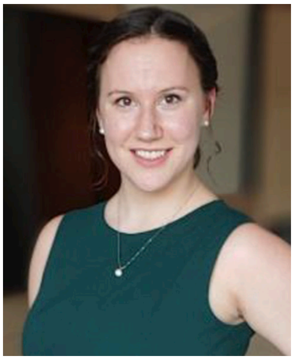
Topping, Keith J, and Stewart W Ehly. 2001. "Peer Assisted Learning: A Framework for Consultation." *Journal of Educational and Psychological Consultation* 12 (2): 113–32.

Van Lehn, KA, MTH Chi, W Baggett, and RC Murray. 1995. "Progress Report: Towards a Theory of Learning During Tutoring." *Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh*.

Van Zundert, Marjo, Dominique Sluijsmans, and Jeroen Van Merriënboer. 2010. "Effective Peer Assessment Processes: Research Findings and Future Directions." *Learning and Instruction* 20 (4): 270–79.

Veenman, Marcel VJ. 2012. "Metacognition in Science Education: Defi Nitions, Constituents, and Their Intricate Relation with Cognition." In *Metacognition in Science Education*, 21–36. Springer.

AUTHORS



Courtney Burris is a PhD candidate in the Department of Industrial and Systems Engineering at the University at Buffalo. She is a Schomburg Fellowship alum and a current Western New York Prosperity Fellow. She earned a B.S. from Elmira College in Mathematics, and a M.S. from the University at Buffalo in Industrial and Systems Engineering. She is working under the guidance of Dr. Alexander Nikolaev in the Social Optimization Lab. Her research interests include engineering education, curricular analytics, sports analytics and social network analysis.

Alexander Nikolaev is an Associate Professor in the Department of Industrial and Systems Engineering at the University at Buffalo. He has a B.Sc. from Moscow Institute of Physics and Technology, and M.S. from the Ohio State University and PhD from University of Illinois at Urbana-Champaign (both in Industrial Engineering). Dr. Nikolaev's interests and expertise are in resource allocation under uncertainty, social network analysis, causal inference, and decision-making that supports pro-health, pro-environmental and educational programs. He has (co-)authored 50s publication in archival journals and referred proceedings, and was a (co-)recipient of University at Buffalo Teaching Innovation Award and INFORMS Impact Prize. He has (co-)authored over 50 manuscripts published in archival journals.

Himangshu Kumar Paul is a presidential fellow and Ph.D. candidate in operations research in the Department of Industrial and Systems Engineering at the State University of New York at Buffalo. He obtained his B.S. and M.S. degrees in Industrial and Production Engineering from Bangladesh University of Engineering and Technology, and has been serving there as lecturer. He is currently working at the Social Optimization Lab under guidance of Dr. Alexander Nikolaev. His research interests include social network modeling, stochastic simulation, and statistical learning.

Lan Peng is a Ph.D. student in the Department of Industrial and Systems Engineering at the University at Buffalo, SUNY. He received a Bachelor's degree in Quality and Reliability Engineering from Beihang University, in 2015, and a Master's Degree in Control Science and Engineering from Beihang University, in 2018. His research interests focus on vehicle routing problems and other combinatorial optimization problems related to the "Last-Mile-Delivery".

APPENDICES

The question ranking phase has two technical components: a task allocation algorithm that assigns questions to students (for comparison in pairs), and a ranking algorithm which determines the aggregate rank of the questions from the results of these pairwise comparisons. As previously discussed, the CRC method can run smoothly with any approach to question assignment and ranking, however, the algorithms can facilitate the practice for an instructor and provide "equal opportunities" to all students.

Question Comparison Assignment

An Integer Programming (IP) model can be used to assign question pairs to students. The objective here is to find such a feasible solution that allows for the highest possible level of exposition of the subject material — number of distinct questions — for each student. Table 2 displays the sets, parameters, and the decision variable used in the question assignment IP model. Table 3: The sets, parameters, and decisions variables for the question assignment integer programming (IP) model.

Table 3. The sets, parameters, and decisions variables for the question assignment integer programming (IP) model.

Sets	
Q	Set of questions
S	Set of students
B	$B \subset Q$, set of 'best' questions
W	$W \subset Q$, set of 'worst' questions
P_s	$P_s \subset Q$, $s \in S$, set of questions provided by student s
Parameters	
b_{ij}	$i, j \in Q$, takes 1 if question i and i forms a ground truth comparison (i.e., question $i \in B, j \in W$ or $i \in W, j \in B$); 0 otherwise
N	Number of comparisons for each student to complete
M	Number of ground truth comparisons for each student to complete
Decision Variables	
x_{sij}	$s \in S$ and $i, j \in Q$, $i < j$, takes 1 if questions i and j are assigned to student s for comparisons; 0 otherwise

The output of the IP model is a set of N pairs of question comparisons for students. To increase students' exposure to material and eliminate bias, students will not be comparing the questions they created. Also, to maximize learner-to-material engagement, students will not compare the same question more than once. In order to access student accountability within the comparing process, each student is assigned M ground truth comparisons. Each set of ground truth comparisons includes a question from the set of "best" questions and a question from the set of "worst" questions. The IP model is as follows.

$$\text{[Dummy] Objective: Maximize 0} \tag{1}$$

s.t

$$x_{sij} = 0 \qquad \forall s \in S, i \in P_s \text{ or } j\ P_s \tag{2}$$

$$\sum_{j \in Q, i<j} x_{sij} + \sum_{j \in Q, i<j} x_{sji} \le 1 \qquad \forall i \in Q, s \in S \tag{3}$$

$$\sum_{i,j \in Q, i<j} x_{sij} = N \qquad \forall s \in S \tag{4}$$

$$\sum_{s \in S,\ j \in Q, i<j} x_{sij} + \sum_{s \in S,\ j \in Q, j<i} x_{sij} \ge \left[\frac{2|S|(N-M)}{|Q| - |B| - |W|} \right] \qquad \forall i \in Q \tag{5}$$

$$\sum_{i,j \in Q, i<j} b_{ij} x_{sij} = M \qquad \forall s \in S \tag{6}$$

Constraints (2) indicate that the student will not compare the questions they created. Constraints (3) ensure that all questions a student compares are unique. Constraints (4) set the number of comparisons assigned to each student, N should be larger than M. Constraints (5) balance the number of times each question gets compared, excluding those from the sets of 'best' and 'worst' questions. Constraints (6) set the number of ground truth comparisons assigned to each student. Note that the objective function is to maximize 0, this problem is a feasibility problem. Multiple equally good solutions may exist for this IP, and its formulation can admit more constraints, if desired.

Ranking Algorithm

The ranking algorithm is implemented during the ranking phase of the CRC process. The completed question comparisons are used as an input for the algorithm and the output is the list of questions with their corresponding rank weight. A random walk based rank aggregation algorithm, called Rank Centrality algorithm, establishes ranking scores of the questions. Rank Centrality has a natural random walk interpretation over the graph of questions, with the edges between those questions that were compared (Negahban, Oh, and Shah 2017). Per the algorithm's logic, the ranking score of a question is its stationary probability for this random walk. Rank Centrality falls into the category of *spectral ranking algorithms* that compute or approximate the leading eigenvector of the random walk transition matrix. The question ranking process consists of four steps: (1) reading the data, (2) cleaning the data, (3) aggregating the data into an integer square matrix, and (4) building the transition probability matrix.

The first step is to read the raw data. Each student performs a set of pairwise comparisons, each of which are randomly selected and the corresponding results are collected. The data should be structured with student IDs as row headers and corresponding comparison responses in the columns. Note the number of columns is equal to the product of the number of students and the number of assigned comparisons and each column should have only one entry. Each response cell contains the following information: the IDs of the questions of the pair and the response provided by a student (e.g., question A is more useful than question B). The response is made on a qualitative Likert scale. The code reads the raw data file and stores the information into a NumPy array.

The next step is to clean the data by converting the qualitative rating (e.g., more useful or way more useful) into a quantitative rating scale (e.g., 1-9). Then, we search for multiple entries (rows) with the same student ID and only keep most complete or recent entry for each student and remove columns with no entries.

The third step is to aggregate the response data into an integer square matrix (of edge size equal to the number of questions) containing aggregated preference score, where the rows represent the losing question and the columns represent the winning question of a pair. Let's consider couple of examples to understand how the preference matrix is created from response data: (1) If the response is "question A is more useful than question B", then 3 points will be added to the cell located at Column A and Row B; and (2) if the response is "question A is way more useful than question C", then 9 points will be added to the cell located at Column A and Row C. Please note that the preference matrix will be initiated with 1 in every cell except the diagonal ones, to ensure an irreducible Markov chain. In order to build this matrix, the code extracts three numbers from each response: ID of the winning question, ID of the losing question, and the preference score. The code then aggregates all the responses and return the preference matrix as a NumPy array.

The fourth step is to build a transition probability matrix from the preference matrix. This step enables us to utilize the random walk based Rank Centrality algorithm. A Transition Probability Matrix is a square matrix with each row summing to 1 which is used to describe the transitions of a Markov chain. The random walk is represented by a transition probability matrix $P = [P_{i,j}]_{nxn}$. $P_{i,j}$ is defined as,

$$P_{i,j} = \begin{cases} \frac{1}{n-1}\left(\frac{r_{i,j}}{l_{i,j}}\right) & i \neq j, \; 1 - \frac{1}{n-1}\sum_{\lambda \neq i} \frac{r_{i,j}}{l_{i,j}} \quad i = j. \end{cases}$$

where, n is the total no. of questions, $r_{i,j}$ is the aggregated winning score of question j over question i and $l_{i,j} = r_{i,j} + r_{j,i}$.

The final step to compute the ranking of the questions is to solve the system of linear equations $\pi^T = \pi^T \cdot P$, where π is the stationary distribution of the random walk and also the ranking score known as "Rank Centrality". Questions with the highest ranking score is considered "most preferred". We create a custom function named "TPM_and_rank" in the code to compute the Transition Probability Matrix from the preference matrix, and subsequently, the ranking score of the questions.