

## Article

# Exploring (Collaborative) Generation and Exploitation of Multiple Choice Questions: Likes as Quality Proxy Metric

Heinrich Söbke 

Bauhaus-Institute for Infrastructure Solutions (b.is), Bauhaus-Universität Weimar, Goetheplatz 7/8,  
99423 Weimar, Germany; heinrich.soebke@uni-weimar.de

**Abstract:** Multiple Choice Questions (MCQs) are an established medium of formal educational contexts. The collaborative generation of MCQs by students follows the perspectives of constructionist and situated learning and is an activity that fosters learning processes. The MCQs generated are—besides the learning processes—further outcomes of collaborative generation processes. Quality MCQs are a valuable resource, so that collaboratively generated quality MCQs might also be exploited in further educational scenarios. However, the quality MCQs first need to be identified from the corpus of all generated MCQs. This article investigates whether Likes distributed by students when answering MCQs are viable as a metric for identifying quality MCQs. Additionally, this study explores whether the process of collaboratively generating MCQs and using the quality MCQs generated in commercial quiz apps is achievable without additional extrinsic motivators. Accordingly, this article describes the results of a two-stage field study. The first stage investigates whether quality MCQs may be identified through collaborative inputs. For this purpose, the Reading Game (RG), a gamified, web-based software aiming at collaborative MCQ generation, is employed as a semester-accompanying learning activity in a bachelor course in Urban Water Management. The reliability of a proxy metric for quality calculated from the ratio of Likes received and appearances in quizzes is compared to the quality estimations of domain experts for selected MCQs. The selection comprised the ten best and the ten worst rated MCQs. Each of the MCQs is rated regarding five dimensions. The results support the assumption that the RG-given quality metric allows identification of well-designed MCQs. In the second stage, MCQs created by RG are provided in a commercial quiz app (QuizUp) in a voluntary educational scenario. Despite the prevailing pressure to learn, neither the motivational effects of RG nor of the app are found in this study to be sufficient for encouraging students to voluntarily use them on a regular basis. Besides confirming that quality MCQs may be generated by collaborative software, it is to be stated that in the collaborative generation of MCQs, Likes may serve as a proxy metric for the quality of the MCQs generated.



**Citation:** Söbke, H. Exploring (Collaborative) Generation and Exploitation of Multiple Choice Questions: Likes as Quality Proxy Metric. *Educ. Sci.* **2022**, *12*, 297. <https://doi.org/10.3390/educsci12050297>

Academic Editor: Mike Joy

Received: 28 February 2022

Accepted: 19 April 2022

Published: 21 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** user-generated content; game-based learning; quiz app; collaborative content generation; MCQ; mobile learning; multiple choice question assessment scheme

## 1. Introduction

Multiple choice questions (MCQs) have been used in educational contexts for a long time, predominantly for assessment purposes (e.g., [1–3]). MCQs are not restricted to a certain domain; they may cover almost all knowledge domains and (at least in theory) all complexity levels of knowledge [4–6]. Besides testing, MCQs may be used for learning; additionally, there is a so-called *testing effect* [7,8]: repeated answering of MCQs leads to memorization—although not all the effective mechanisms are well-known so far [9]. The testing effect is not specific to MCQs, but to testing in general. Larsen & Butler (2013) [10] subsume MCQs to recognition tests. In contrast, fill-in-the-blank-tests—as, for example, investigated by Wiklund-Hörnqvist, Jonsson, & Nyberg (2013) [11]—belong to production tests. Retrieval from memory is considered as a fundamental cognitive process producing this effect [12]. An overview of work in the context of repeated tests is presented by

Roediger & Karpicke (2006) [8]. Furthermore, effects on long-term retention [13], the impact of type and timing of feedback [14], the influence of frequency of quizzes and their placement relatively to corresponding lectures [15] are discussed in the literature. According to Glass, Ingate, & Sinha (2013), better long-term retention occurs if an MCQ has been included in a final exam [16]. Additionally, the importance of quizzing for learning is underlined by the recommendation of the IES National Center for Education Research (U.S. Department Of Education) [17].

These insights have led to a lot of MCQ-based learning tools. Quizlet [18], Quitch [19], and KEEUNIT Quiz-App [20] are only a few, which are established commercial products. By the provision of mobile clients, they support ubiquitous accessibility and, in consequence, mobile learning, regardless of time and location [21].

On the other hand, quizzes, which are based on MCQs, are an established entertainment format, which has demonstrated its attractiveness in many contexts: pub quizzes; quiz shows on TV such as *Who Wants to Be a Millionaire?* [22]; quiz board games, e.g., *Trivial Pursuit* [23] and, since 2013, quiz apps such as QuizClash [24] and QuizUp [25]; and many other instantiations of quizzes have received a huge popularity. A few commercial quiz apps, among them QuizClash and QuizUp, have been installed millions of times [26]. Thus, quizzes provide attractive gaming experiences. Such gaming experiences are also considered as fostering motivation for learning [27,28]. In particular, quiz apps seem to have excellent potential as learning tools, as users are eager to learn and perceive competition within their social networks as motivating [29].

A frequent environment for quiz applications are classroom settings [30], which are supported, for example, by Kahoot! [31], Socrative [32] or arsnova.click [33,34]. The motivational effects of these synchronous classroom settings are well documented, for example for Kahoot! [35–38] or for Socrative [39–41]. Another application area is asynchronous educational scenarios, which are location- and time-independent and often are based on apps using mobile devices [42,43], such as the commercial quiz apps QuizClash and QuizUp, described above. However, in educational scenarios, commercial quiz apps are not used, but rather specific quiz apps specifically developed for learning, such as Quitch [19]. A German financial service provider uses such a quiz app [20] for product training [44]. Although great success has been achieved using specific quiz apps (e.g., [42,43]), the question arises to what extent commercial quiz apps, which are optimized for entertainment and thus motivation, may be used as learning tools as well. No studies have been found examining whether commercial quiz apps designed for entertainment deliver sufficient motivation for students to use these quiz apps in voluntary learning activities. To this end, however, a commercial quiz app needs to be capable of accommodating subject-specific, user-defined MCQs. Amongst others, the commercial quiz app QuizUp, which was opened up to user-defined MCQs a few years after its release [45], is such an app. In particular, QuizUp also qualifies as a learning tool because matches do not cover multiple topics, but only one topic, allowing the game to focus on the topic to be learned.

Another prerequisite for the adoption of quiz apps is the availability of quality MCQs, characterized, for example, by distractors that would represent plausible choices (e.g., [5]). The generation of quality MCQs is not trivial and takes—if done manually by domain experts—a lot of effort. Besides automated generation (e.g., [46]), crowdsourcing [47,48] is an attempt to mitigate these efforts: assigning the task of MCQ generation to learners not only distributes the efforts but is also considered as beneficial for their knowledge acquisition and skill development from the perspective of constructionism [49]. Campbell (2010) found, in a study using the learning management system moodle [50], that the creation of questions stimulated both weaker and stronger learners to learning [51].

Jones (2019) [52] revealed that generating MCQs is perceived by students as a valuable learning activity and that students tend to prefer collaborative activities for the generation of MCQs. Overall, generating MCQs is seen as conducive to learning [53], for example, in the context of active learning approaches [54]. Furthermore, creating MCQs is also seen as a good preparation for MCQ tests themselves [55]. In particular, giving feedback

promotes learning [56]. Student-generated MCQs may be used in a manual way, for example, being discussed in the classroom after being polished by the lecturer and made available in discussion forums [57]. Yet, web platforms for collaborative question generation are also available, such as PeerWise [58]. Positive effects on learning processes include increasing learner engagement and fostering collaborative learning [58–60]. However, challenges emerge when students consider writing MCQs difficult without triggering adequate subjective learning gains at the same time. In addition, students do not trust the questions of their peers [61]. Additionally, the concealment of the identity of students in collaborative questioning is considered beneficial [62]. Despite all challenges, systems based on user-generated MCQs—such as Quizzical [63], RiPPLE [64], or UpGrade [65]—are considered to have a great learning potential, provided they are supported by external incentives in formal educational scenarios, as was done in the referenced studies.

This article complements the fundamentals described with the results of a two-staged field study. The objectives of the study were to investigate collaborative MCQ generation and the identification of quality MCQs in the first phase and the use of the collaborative question generation software and a commercial quiz app for further utilization of MCQs to voluntary learning scenarios in the second phase. In a pragmatic approach, the MCQs created in the first phase may be used in a motivating learning activity in the second phase. The research questions to be answered are: for the first stage (RQ 1), to what extent the quality of the generated MCQs may be measured by collaboratively generated data and for the second part (RQ 2) whether the collaborative generation of MCQs as a process is sufficiently motivating that the process works without extrinsic incentives of educational scenarios and—if the process does not work without incentives of the educational scenario—whether instead a commercial quiz app succeeds without such incentives of educational scenarios.

In the following section, the methodology is described; thereafter, the results of the first step (RG as tool to generate MCQs) and the second step (QuizUp as a tool to answer MCQs) are described. Thereafter, the results are discussed, and, finally, conclusions are derived.

## 2. Materials and Methods

Two software tools were used: in the first stage, MCQs are created and ranked by students. In the second stage, these MCQs are provided for learning purposes via a well-established commercial quiz app. In the following, employed software tools and the experimental settings are described.

### 2.1. Generating MCQs: The Reading Game

The Reading Game (RG) is implemented as a moodle [50] module. It aims at prompting users to generate quality MCQs. During the study, RG was still in an experimental state. RG is an epistemic game that requires students to create questions and answer questions at a predefined regular rate, indicated by a control bar [66]. Because students may review MCQs by adding comments as well as by liking them or by reporting them to an administrator, this gamified application provides a form of collaborative question engineering.

### 2.2. Answering MCQs: QuizUp

Generated MCQs have been provided in a commercial quiz app. In this way, the vast popularity and attractiveness of commercial quiz apps should be exploited for learning. In September 2015, QuizUp, a major commercial quiz app, was opened for user-defined content [45]. This opening has created an opportunity to add user-defined topics to QuizUp sourced from RG mechanisms.

### 2.3. The Experimental Setting

In the first step, RG was introduced to a bachelor course in Urban Wastewater Management ( $n = 16$  (students, who took the final exam) resp.  $n = 29$  (students, who enrolled in the course initially)). Students were instructed by a written assignment description to fulfill

their weekly quota, which requires them to create one MCQ and to answer five MCQs of their fellow students each week. Students have to meet the weekly quota in 10 of 16 weeks of the semester in order to be admitted to the final exam. Besides this formal incentive, students were encouraged in the written assignment description to use the Like feature to identify any well-generated MCQ. Students were also made aware of the Comment feature for contributing to the improvement of MCQs. Overall, students were alerted that engaging with MCQs would support the learning objectives of the course. An administrator ensured that reported MCQs (MCQs that have been marked as wrong by the participants) were either approved or deactivated.

In the second stage, a group of five students used another instance of the RG in the context of the bachelor course Capital Budgeting (CB). As part of a project assignment, the five students were mandated to investigate the impact of RG. First, the five students generated a basic corpus of MCQs through collaborative playing of RG for three weeks at the beginning of the semester. After these three weeks, the students began encouraging their fellow students ( $n = 30$ ) from the CB course to participate in RG as well, so that all fellow students could prepare for the final exam. After none of the fellow students participated in RG, and feedback indicated that fellow students shied away from generating MCQs, the instructional design was adjusted: Instead of answering the MCQs in RG, the MCQs generated were transferred to a topic Capital Budgeting in QuizUp, where the MCQs were available for practice. The response for this topic in QuizUp was again poor. To evaluate the reasons for student inactivity, the group of students employed a self-designed questionnaire, which was answered by 18 fellow students.

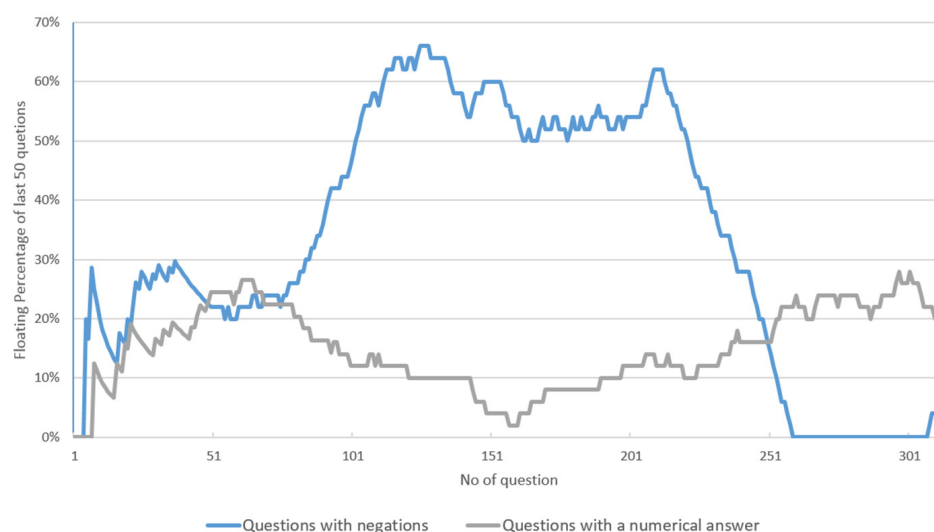
### 3. Results

#### 3.1. Stage 1: MCQ Generation Using the Reading Game (RG)

##### 3.1.1. Usage Data

In the 16 weeks during which the game was active during the semester term, 29 users created 379 MCQs and answered 6689 MCQs. 326 MCQs were liked, i.e., Likes were issued at a low rate of only 0.5%. 15 MCQs were reported by students. All MCQs reported were finally deactivated. The comment feature of the RG was used very seldom, mainly for communication about reported MCQs. There was only one game-related student inquiry about the game—a clarification about an MCQ deactivated. Four of the 29 students did not adhere to their weekly quota and were not admitted to the final test.

A first observation was that prevalently negated MCQs were generated after an orientation phase for the students. Negated MCQs are quite easy to generate, as they free the originator from the demanding task of finding well-balanced distractors. Instead of three distractors, only one distractor has to be found; the other three answer options are formed by correct answers. By use of negation, this distractor becomes the correct answer. However, this approach is not recommended as a good practice to build MCQs [67]. Figure 1 illustrates that after approximately a third of the semester, students predominantly generated negated MCQs. More than half of all MCQs generated included a negation for almost a third of the course period. This phenomenon was stopped after the course administration instructed students to avoid this type of MCQ. Soon after this instruction, the percentage of MCQs asking for a numerical answer increased again, doubling from 10% to 20%. Besides negation, an MCQ asking for a numerical answer is another type of MCQ, which eases MCQ generation because distractors (other numerical values) might be found with little effort. Although MCQs asking for numerical answers are also to be regarded as MCQs of high quality, focusing solely on MCQs with numerical answers underrepresent specific domains as numerical answers that are more common in specific domains, such as mathematics or physics [68]. Further, not all learning objectives are supportable by MCQs employing numerical answers.



**Figure 1.** Floating percentages of MCQ generated showing negations and numerical answers.

A further observation was that throughout the game a small group of 3–4 students battled for the leading position in the game and they answered up to 10 times as many questions as required for admission to the final exam. This phenomenon has been observed in other competitions including educational quizzes as well [69]. Another group of students fulfilled just their weekly quota referring to both MCQs the generation of MCQs to answer. In between, there was a group which did not stick immediately to the quota but answered an extra number of MCQs from time to time. A classification might be hypothesized into competitive overachievers, interest-driven casual learners, and effort-optimizing minimalists. The group of effort-optimizing minimalists may be interpreted as evidence suggesting weekly quotas as an admission requirement for the final exam acted as a trigger to start participation in RG. In general, the number of MCQs answered varied much more than the number of MCQs generated. On average, a student answered almost 500% of the mandatory quota, but only generated 40% more MCQs than required. These ratios may indicate that generating MCQs is perceived to be significantly more difficult than answering MCQs. Table 1 includes the minimum, maximum, and average number of MCQs answered and generated. In addition, the number of MCQs required to be admitted to the final exam is given as an indicator value. The numbers refer to those 16 participants who completed the final exam and therefore should be considered as the most regular users of RG.

**Table 1.** Number of MCQs answered and generated ( $n = 16$ ).

	Required	Min	Max	Mean
MCQs answered	50	77	502	239.9
MCQs generated	10	11	17	14.3

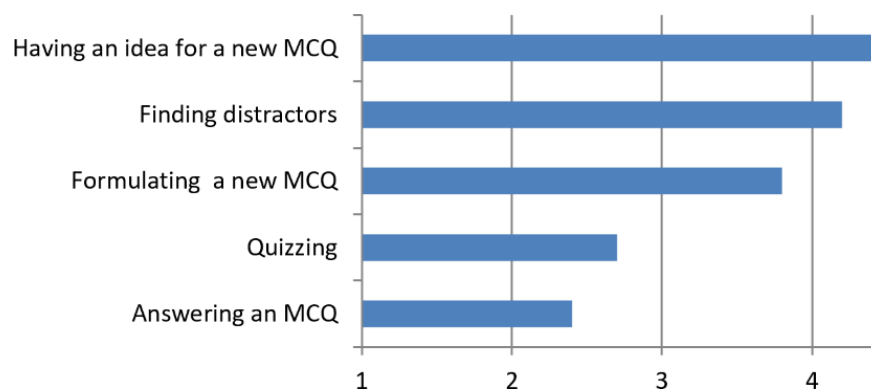
### 3.1.2. Students' Perceptions

Upon completion of the RG activity, however, before the final exam, students were asked to answer a questionnaire consisting of 21 questions. 26 answers were received; 10 of the respondents decided either not to take the final exam or were not admitted to the final exam. 19 (73%) of them took part regularly in RG because they wanted to receive admission to the final exam. The other 5 participants (19%), who were already admitted to the test in a previous semester, wanted to prepare for the final exam. The majority of participants (73%) logged into RG once a week. 50% of respondents estimated the weekly time spent in RG as 10 to 20 min.

Respondents were asked for their estimation of the difficulty of various tasks in RG. Answers on a 6-point Likert scale are depicted in Figure 2. The effort of creating a new MCQ

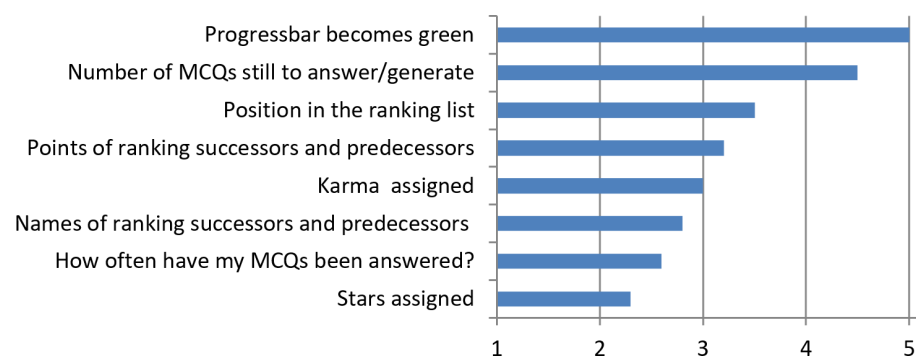


was marked as very challenging. All related tasks (having an idea, finding distractors, and formulating the MCQ) received higher values of perceived difficulty than the alternatives of quizzing (answering 10 MCQs in a row) and answering a single MCQ. Noteworthy is the huge difference between both categories of difficulties: tasks related to creating MCQs are rated almost 2 points more difficult than just answering MCQs.



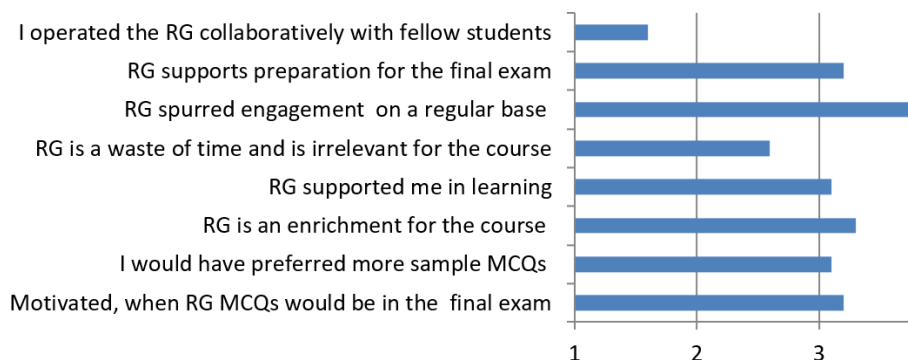
**Figure 2.** Perceived task difficulty; 6-point Likert scale from 1 (very easy) to 6 (extremely difficult) ( $n = 26$ ).

A further question is the efficacy of gamification elements. RG is positioned as a kind of game. Hence, a fundamental question is the expectation with which a student enters RG: are they in the mood to play a game or to use a learning tool? Respondents were asked for the main information elements, which they considered as important for the operation of the RG. The results of this question, which again used a 6-point Likert scale for each control element, are depicted in Figure 3. The most-observed information is the control bar. This might be not surprising, as this gauge is the official measurement that the course administrators stick to. Related to this indicator are the given numbers of MCQs still to be answered and asked. The most important gamification element is competition, as the next indicator—the position in the ranking list—suggests. Remarkable here is the large gap of more than one point relative to the previous indicators. The urge towards competition is assisted by the information regarding how many points are needed to move one rank up. Competition in this context seems not to be strongly personal, as names of the ranking list neighbors are not that important. Assigned karma, i.e., a measurement of how much a person's MCQs are liked and therefore a kind of recognition from fellow students for one's work, seems at least to be noted, together with the information of how often one's MCQs generated have been answered. The most ignored kind of information is the stars assigned. This is a weekly reward for most points and most answered MCQs, an achievement, which is considered as a classic gamification element [70].



**Figure 3.** Relevant control information for RG: 6-point Likert scale from 1 (I do not know) to 6 (Main control information) ( $n = 26$ ).

The next group of questions evaluated students' perception of RG as a learning tool. As Figure 4 reveals, there are no settled statements. The only denied statement is that RG is operated collaboratively. Further, the respondents admitted that RG stimulated learning activities on a regular base. While respondents were undecided, if the game supported them to learn, they mostly rejected the (concededly provocative) statement that RG is a waste of time.



**Figure 4.** Approval of statements; 6-point Likert scale from 1 (not at all) to 6 (yes, exactly) ( $n = 26$ ).

In general, RG was received more as additional work than as a game. In a comparative question about preferred lecture-accompanying learning activities, RG got the lowest marks (2.7) on a 6-point Likert scale compared to online questionnaires (4.3) and calculation exercises (3.2). Especially, generation of MCQs led to avoidance behavior (negated MCQs).

### 3.1.3. Analysis

The data described above were linked to further data from the didactic scenario, described in Table 2. Students not only completed the RG during the semester, but also had to complete regular online pretests as a further requirement for admission to the final exam. Each of the 9 online pretests consisted of 5 MCQs. 7 of them had to be passed with at least 60% to obtain admission to the final exam. For these 9 tests, a pool of 140 MCQs was used from previous semesters of the course. This pool has been enlarged by 32 MCQs from RG, which have been identified by the number of Likes received. The RG MCQs were checked for technical accuracy by two domain experts and adapted accordingly. A selection of MCQs from this pool was also the subject of the final exam [71].

**Table 2.** Data collected per student.

Data (Variable)	Description
MCQs answered (A)	Number of MCQ answered in RG
MCQs generated (B)	Number of the MCQs generated in RG
Points (C)	Points in RG, this variable is derived from variable A and B.
Performance in online pretests (D)	Percentage of correctly answered MCQs
No. of mock tests completed (E)	Between last lecture of the course and final exam, students could train the MCQs of the MCQ pool by means of a mock test, each consisting of 5 randomly selected MCQs.
Final exam results (MCQs) (F)	In the final exam, students had to answer MCQs, which have been issued in accordance with those of the MCQ pool.
Final exam result (Calculations) (G)	In the second part of the final test, students had to solve calculation tasks.

This data ( $n = 16$ ) was subjected to a correlation analysis. All absolute values found to be higher than 0.4 are included in Table 3. The final exam results (MCQs) (F) are positively correlated with the number of mock tests completed (E), the number of MCQs answered (A) in RG, and the points (C) in RG. As C and A may be regarded as an indicator of time spent for test preparation, the values for the correlation coefficients seem to be reasonable.

The increased value of 0.79 between the sum of A and E and the final exam results (MCQs) (F) seems to be reasonable, as efforts of mock tests and answering MCQs may be mutually substitutable. Overall, the correlations shown are in line with the recognized assumption that active engagement with the learning subject matter results in better learning outcomes. [72,73].

**Table 3.** Correlation coefficients of variables in RG scenario.

Variable 1	Variable 2	Correlation Coefficient
F (Final test results (MCQs))	E (No. of completed mock tests)	0.68
F (Final test results (MCQs))	A (MCQs answered)	0.48
F (Final test results (MCQs))	C (Points)	0.53
F (Final test results (MCQs))	A (MCQs answered) + E (No. of completed mock tests)	0.79
A (MCQs answered)	B (MCQs generated)	0.58
G (Final exam result (Calculations))	B (MCQs generated)	−0.60

An unexpected result is the negative correlation of  $-0.60$  between MCQs generated (B) and the final exam result (Calculations) (G). Whether participation in RG was motivated by compensation, especially among students who had weaknesses in the calculation tasks to be solved in the final exam, still needs to be explored. In a previous study, a few students reported that engaging in quizzes for a short period of time led, in particular, to a sense of relief from having contributed towards learning [71].

#### 3.1.4. Analysis of the MCQs Generated

A special kind of feedback, which students may give in RG, is liking an MCQ answered. The number of Likes for all MCQs of a participant are aggregated: As Karmascore, the Likes are a form of reward for well-received MCQs. The question arises of whether the number of Likes might be used as a measurement for the quality of an MCQ. Especially, MCQs generated by students need to be assessed for their quality, also because students have doubts about the quality of self-generated MCQs [52]. Besides peer assessment of quality, a further approach to assessing the quality of MCQs is the assessment by experts, who examine the MCQs for supported educational goals; for example, in [74] an MCQ pool is mapped to Bloom's taxonomy of educational goals [75]. Artificial intelligence may also be used. [76].

To determine whether Likes are a valid proxy metric for selecting quality MCQs, expert judgment was used as a reference here. A number of MCQs and an assessment scheme for MCQs were included in a questionnaire. This questionnaire was answered by domain experts. Finally, it was evaluated whether there are correlations between the number of Likes an MCQ has received and the assessments of the domain experts. In the following, the steps of the methodology are described:

**Selection of MCQs.** Both 10 well-rated and 10 not-so-well-rated MCQs from the RG-corpus of 379 questions were selected to provide a broad range of quality. Those with at least three Likes and the best ratio of Likes per answer (Karmascore) were selected as the best MCQs. The ten worst rated MCQs were identified as those with the most answers without any Likes. Finally, the selected 20 MCQs were included in an arbitrary order in a questionnaire, so that the quality could not be inferred from the position in the questionnaire.

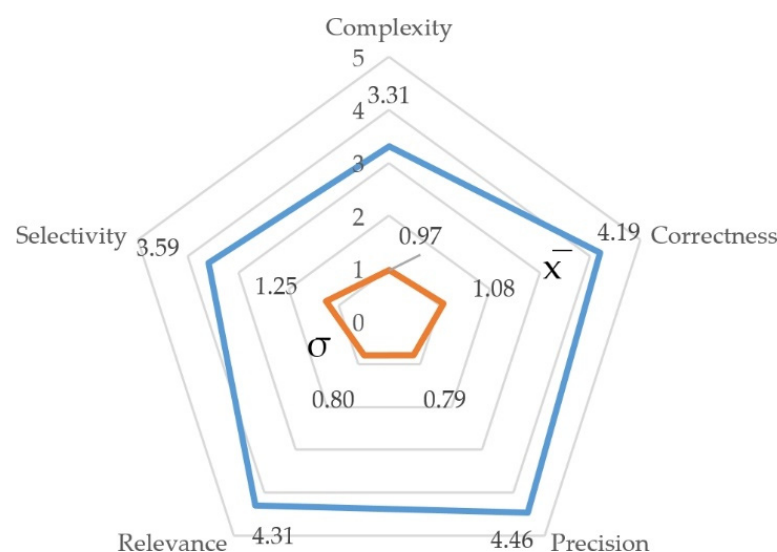
**Assessment scheme.** An assessment scheme for MCQs was developed guided by the work of Haladyna & Rodriguez (2013) [67]. The dimensions of the scheme are presented in Table 4. Each MCQ has to be rated according to each dimension. A 5-point scale from 1 (not at all) to 5 (yes, completely) has been used.



**Table 4.** Assessment scheme for MCQs.

Dimension	Description Including Guiding Questions for the Assessment
Precision	<p>The MCQ is precise and comprehensible.</p> <p>Questions:</p> <ul style="list-style-type: none"> <li>Is the MCQ precise?</li> <li>Is it comprehensible?</li> <li>Is it grammatically and orthographically correct?</li> </ul>
Correctness	<p>Question stem, correct answers and distractors are correct technically.</p> <p>Questions:</p> <ul style="list-style-type: none"> <li>Are Question stem and answers correct from a technically point of view?</li> <li>Are the distractors incorrect?</li> </ul>
Relevance	<p>The MCQ's content is relevant for the technical domain.</p> <p>Question:</p> <ul style="list-style-type: none"> <li>How important is the content to capture the knowledge of the subject area?</li> </ul>
Complexity	<p>The knowledge given by the MCQ is complex.</p> <p>Questions:</p> <ul style="list-style-type: none"> <li>What is the complexity level of the MCQ's knowledge?</li> <li>Is it factual knowledge, procedural knowledge, or system knowledge (increasing complexity)?</li> <li>How difficult is it to answer the MCQ correctly?</li> </ul>
Selectivity	<p>The distractors are well selected.</p> <p>Questions:</p> <ul style="list-style-type: none"> <li>Are distractors chosen in a way which requires knowledge of the subject area?</li> <li>Is the correct answer selectable without any knowledge of the subject area? (Negative)</li> </ul>

**Expert survey.** To obtain a rating for the 20 MCQs regarding all 5 dimensions, domain experts from the chair for Urban Wastewater Management and the affiliated institute were invited to evaluate the MCQs. In summary, the questionnaire was answered by 18 domain experts. Figure 5 depicts mean value and standard deviation for each dimension. Most deficits show the dimensions Complexity and Selectivity, whereas Precision, Correctness, and Relevance received rather high values.

**Figure 5.** Expert assessment (mean value  $\bar{x}$  and standard deviation  $\sigma$ ) of 20 MCQs regarding 5 dimensions (5-point Likert scale,  $n = 18$ ).

**Results.** For each MCQ, the dimensions' mean values and the Karmascore have been analyzed for correlations (see Table 5). The best value of 0.51 for a correlation has been found between Relevance and Karmascore. Complexity follows with a value of 0.34, whereas *Precision* and *Correctness* show rather low values and *Selectivity* seems not to be correlated to Karmascore at all.

**Table 5.** Correlation coefficients  $r$  of expert evaluations and Karmascore.

Dimension	$r$ Dimension Karmascore
Complexity	0.34
Correctness	0.25
Precision	0.17
Relevance	0.51
Selectivity	−0.02

As a summary, MCQs that receive Likes by students seem to be characterized mostly by Relevance and Complexity. Therefore, the Like feature is to be considered as valid for ranking MCQs according to their quality.

### 3.2. Stage 2: The Reading Game and QuizUp in a Capital Budgeting Course

#### 3.2.1. Motivation and Method

In the previous setting, participation was mandatory for all students, as there were means to sanction their non-participation. At least in theory, RG provides a frame that may be filled by the self-directed and self-paced work of the students, and that provides relatively short-cycled feedback by statistics and by Karma, provided by fellow students. Therefore, it is worthwhile to test whether RG fosters motivation sufficiently by letting students participate in the game voluntarily. The research question was whether RG can serve as a tool in an informal learning context, i.e., without a formal obligation.

**The Starting Phase.** The study was started as a student project of five participants in their studies for a bachelor's degree. In parallel, they had to take part in the course Capital Budgeting (CB). Their task was to operate an instance of RG. Thereby they should build up a pool of MCQs. For growing the number of MCQs quickly, each member of the team had to provide three MCQs in the first two weeks. Participants claimed that CB would not be an appropriate domain for MCQs: there would be only little knowledge to memorize, but mostly just procedural knowledge would have to be applied. Consequently, a further MCQ type was introduced: rough estimation MCQs. These MCQs are to be solved by mental calculations; they should transform procedural calculation knowledge into MCQs. The low number of participants became a problem in later weeks, when new MCQs were not generated sufficiently, and participants could not fulfill their quota without interventions.

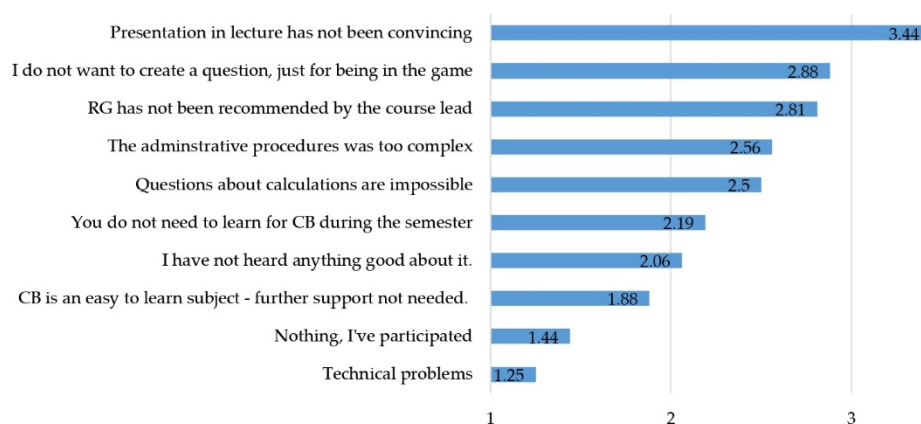
**The Blossoming Phase.** After three weeks, the project group advertised RG in a short introduction in the lecture with the intention of inviting their fellow students to join. The projected group repeated this invitation two times. No other student joined the game. As the project group indicated from personal conversations with their fellow students, the main worry of their peers was about the requirement to generate an MCQ. The fellow students stated a desire to benefit from RG by answering the MCQs but considered the effort of creating MCQs as too much. Thus, there was no blossoming phase.

**The Harvesting Period.** According to the preferences of the fellow students to have access to the MCQs without the obligation to create MCQs, the project group transferred the MCQs to the commercial quiz app QuizUp. Thus, the user-defined Quiz Up topic Capital Budgeting, including 57 MCQs, was created. Remarkably, in the project session, where the project group were first introduced by their advisors to the option of transferring their RG MCQs to QuizUp, the project group unintentionally demonstrated the low-threshold accessibility of mobile apps by all taking their mobile device unrequested and installing the app within five minutes.

Again, the new QuizUp topic was advertised in a lecture by the students. The result was disappointing again; only six (out of 30) students tested the topic but did not use it regularly. Altogether, QuizUp seems not to be an attractive tool in informal educational contexts. Consequently, the reasons for not participating either in RG or in playing QuizUp—though the provided contents were relevant for the written test—were collected by a questionnaire.

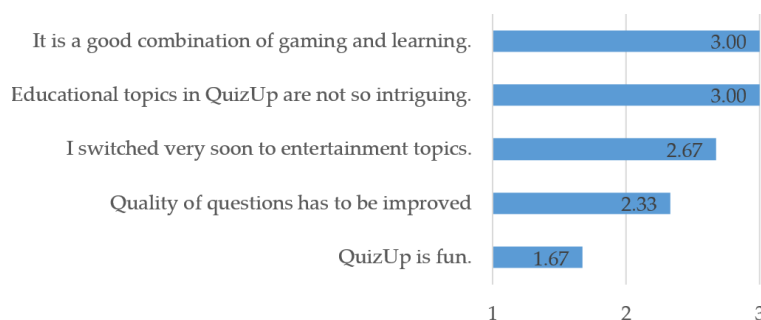
### 3.2.2. Questionnaire

The questionnaire consisted of 9 questions in the categories RG, QuizUp and Learning (see: Supplementary Materials). It was launched after the last lecture supported by the lecturer of the course CB. 20 answers were received; 15 of them completed the questionnaire. The first question asked whether respondents were aware of RG. 18 of 20 confirmed. A second question asked for the reason not to enter RG (Figure 6). Again, participants could indicate their reasons on a 5-point Likert scale. The statement rated highest indicated that the presentation in the lecture was not convincing, meaning that students could not envision a beneficial learning situation. Together with the second most named reason, the unwillingness to create a question and the lack of formal approval for this tool, these answers might serve as an explanation for missing participation in RG. Further hindrances were mistrust of the idea that such a game might contribute to learning and that semester-accompanying learning is useful and required.



**Figure 6.** Reasons not to participate in RG (5-point Likert scale,  $n = 18$ ).

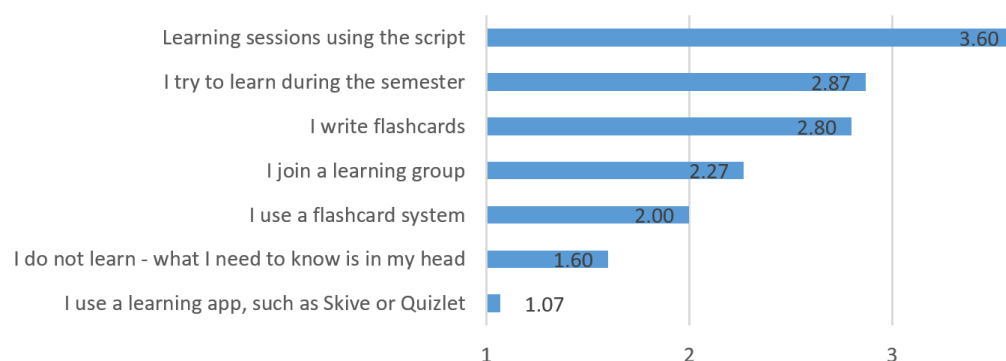
The propagation of QuizUp and its educational topic CB was not successful. Only 4 of 17 respondents were aware of QuizUp, and only 3 of them already had experiences in QuizUp. Figure 7 summarizes a list of the non-representative attitudes of only 3 participants. Nevertheless, new questions arise about the use of QuizUp: Is playing QuizUp fun (as suggested by its commercial success)? Is there a difference between educational and entertainment topics (as suggested in [77]). Finally, is the quality of the MCQs sufficient—which might be an important aspect of accepting quiz apps as learning tools?



**Figure 7.** Attitude towards QuizUp ( $n = 3$ , 5-point Likert scale).

Three verbal responses were received, which pointed to prevalent issues here. The first person indicated that the effort required, in combination with the experimental character, has stopped her from taking part in QuizUp: “I prefer investing my time in learning activities which have already proven their efficiency.” Another person doubts the quality of student-provided MCQs and, further, is not convinced that the complex content of higher education lectures can be transferred into MCQs, a phenomenon which has been described before [61]. Additionally, a third person raises an issue, which applies to a small group of students: visual learning. Instead of dealing with the meaning of the answers, students memorize the visual form of the answers and they identify the correct answer by the length of the words and their visual appearance.

Another question addressed how students approach their learning tasks in general (Figure 8). The most prevalent method is the use of lecture notes for learning sessions. At this point, four written answers indicated that students write summaries of their lecture notes. Learning activities during the semester do not seem to be very popular. A similar observation that students tend to study little during the semester and instead try to prepare for the final exam intensively was also found in an earlier study [71]. Working with flash cards is not too much favored, whereas the use of flash card learning apps is not popular at all. Overall, using digital tools in self-initiated learning activities appeared to be uncommon among the students in this cohort. However, possible reasons that still need to be investigated could be a high proportion of procedural knowledge that may be less well practiced with the help of MCQs but that is, nonetheless, required for the calculation tasks in the final exam.



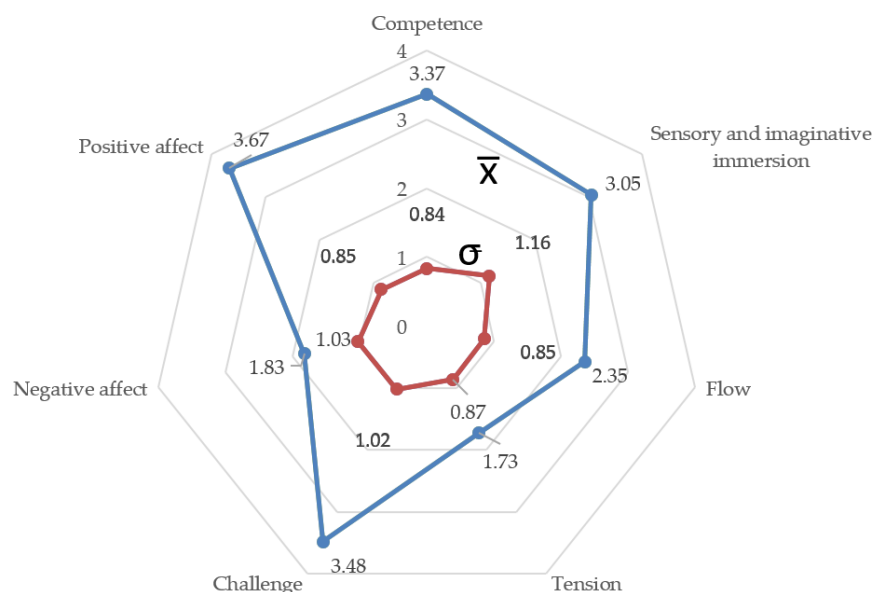
**Figure 8.** Current approaches to memorize factual knowledge ( $n = 20$ , 5-point Likert scale).

#### 4. Discussion

In general, the idea of generating MCQs by a collaborative game for use in further educational scenarios appears to be beneficial to learning. RG-generated MCQs were included in QuizUp and were available for playing. Learning effects have been observed, when students indicated that regular course-accompanying learning activities were triggered by QuizUp. The results are partially in alignment with findings in the literature indicating the positive impact of engagement in generating MCQs on performance in the final exam. [53,58,63]. However, some studies have not observed an impact of engagement in MCQ generation on final exam performance. [78,79].

The results presented appear to be inconsistent with those of an earlier study, in which QuizUp experienced much higher uptake, but in which QuizUp activities were not voluntary [77]. QuizUp has been received as a game, even when it has been used for educational purposes, as the results of the Game Experience Questionnaire (GEQ) [80] suggest. As a learning tool in synchronous lecture settings, it has been accepted. However, as an asynchronous spare time activity, it has received lower acceptance. There were comparatively high values for Positive affect, Challenge, and Competence, whereas values for Negative affect and Tension were low, seemingly typical for a game experience (Figure 9). Further, noteworthy from the previous study is the categorization of players into Learners and Gamers: Learners fulfill their quota of assigned tasks and probably play

a few further entertainment matches, but then leave the app. Gamers, however, accomplish their educational tasks in the game, and then get stuck in the app, i.e., they play 10 times more matches in entertainment topics than in educational topics.



**Figure 9.** Subscale scores (mean value  $\bar{x}$  and standard deviation  $\sigma$ ) of the in-game GEQ (5-point Likert scale,  $n = 26$ , adapted from [77]).

In general, a few limitations of this study need still to be resolved. Certain sample sizes (especially 16 to 29 students operating RG and 18 out of 30 students answering the questionnaire regarding QuizUp) in the study are to be increased in replication studies for attaining greater significance. Further, RG was received more as an assignment instead of a game, although the students valued RG (and QuizUp) as an enrichment of the courses. Only a minor fraction of the students seemed to be susceptible to gaming features. Students accomplished their weekly quota but generating MCQs seemed not to be a preferred task. Additionally, tasks such as liking and commenting were done only reluctantly; thus, the collaborative part of the game did not work as intended. These deficits certainly impacted the quality of the MCQs generated, although the quality has been rated as acceptable, and especially high-quality MCQs might be selected based on the Likes received. In the following, measures aiming at further developments are summarized: organizational measures (including didactical necessities) and software improvements (including game design).

**Organizational Measures.** Didactically, extending the introduction into RG and underlining the positive effects for students might be beneficial for the students' motivation. Further, providing a corpus of well-designed sample MCQs and a design guide for MCQs could provide more orientation to students. Additionally, reviews and assistance by domain experts during the game might improve the learning process and shorten periods of unsound MCQ generation strategies, such as negated MCQs. In general, these experiments confirmed that learning tools such as RG and Quiz Up require a dedicated didactic scenario. Learning activities were not performed voluntarily but had to be spurred by an educational scenario linking learning activities formally to intended course outcomes.

**RG Software Improvements.** The RG module was a prototype and had functional limitations that also reduced effectiveness. For example, commenting on MCQs is not shared with interested students who cannot respond to comments in turn. Thus, discussing a question was cumbersome. In addition, an editing feature was missing. Faulty MCQs had to be deactivated and re-submitted. Additionally, although students were instructed to like an MCQ, only a half percent of all answers was accompanied by a like. However, as shown, Likes may be used to rank MCQs and are therefore important for identifying quality MCQs. Extension of the like feature is suggested in two ways: Firstly, a multi-star rating might



help to clarify the value of a like. Secondly, Likes should be a partially mandatory part of answering an MCQ: when a student has not issued a like for a certain number of answered MCQs in a row, such a rating would appear and would have to be completed. Further options for enhancing the quality of interactions in the game might be introduced, such as mandatory interactions such as assessment or direct competition between participants as a means to contribute to a more meaningful game experience.

**QuizUp.** The usage of commercial quiz apps, such as Quiz Up, in educational scenarios, is not well-known in the literature. A threat of using commercial software in educational scenarios is always the loss of the software, be it due to licensing reasons or due to the discontinuation of the software, as is the case with QuizUp, which was discontinued in the year 2021 [81]. However, a loss may also happen to dedicated software such as RG, which has meanwhile also been discontinued, too. If software is discontinued, there is usually alternative software available that may be used with a one-time setup effort. Related to this study, for example, RG might be substituted by PeerWise [82], along with appropriate instructions for the quotas to be produced. For the substitution of QuizUp, various alternatives are also available, for example the Keeunit quiz app [20]. Further research is required into increasing the attractiveness of educational topics. Further, user-defined topics suffer from some restrictions, which impact the game enjoyment; for example, players are not awarded specific titles when they reach a certain level. Additionally, the behavior of bots as opponents is too simple, e.g., it is almost impossible to beat certain bots, which frustrates players. Among the positive aspects of QuizUp is its openness to all technical domains; thus, it is a domain-independent generic learning tool.

## 5. Conclusions

This two-stage field study investigated two multiple choice question (MCQ)-based digital tools in learning scenarios of bachelor's courses. The first (RQ 1) was the Reading Game (RG), a platform for collaborative generation of MCQs. Regarding RG, it was shown that quality MCQs may be identified by the Likes given by students to their fellow students' MCQs. In the second stage (RQ 2), it was confirmed—in line with findings from the literature—that the process of collaborative generation of MCQs is not solely motivated by the learning outcomes achievable, but must rely on external incentives of framing educational scenarios. Further, MCQs generated by RG may be transferred to a quiz app with little effort from the lecturer, so that the MCQs are available there for further learning scenarios. The commercial quiz app used in this study was the well-established entertainment app QuizUp. Due to the non-mandatory application, the student use of this app was only marginal, too, indicating that even the motivational effects of a commercial quiz app are not sufficient, despite an upcoming final exam, to draw students into voluntary learning activities. Further, both RG and QuizUp have been discontinued in the meantime. Nevertheless, since both learning tools represent a group of learning tools and may be substituted, the results confirm that using collaborative MCQ platforms for generating MCQs, a selection of which is then used in a quiz app in other learning scenarios, is a sustainable and especially domain-independent approach. Furthermore, the results suggest that Likes provide a proxy metric for the quality of collaboratively generated MCQs.

**Supplementary Materials:** The following questionnaires are available online at <https://www.mdpi.com/article/10.3390/educsci12050297/s1>. For each stage of the study, data was collected by a questionnaire, named QuestionnaireStage1.pdf and QuestionnaireStage2.pdf.

**Funding:** This research was funded by the German Federal Ministry of Education and Research (BMBF) through grant 033W011B.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study, due to surveying existing learning scenarios in the field, i.e., the measurements reported here had no influence on the design of the learning scenarios.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Questionnaires used are attached and data is available on request.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- Burton, R.F. Multiple-Choice and True/False Tests: Myths and Misapprehensions. *Assess. Eval. High. Educ.* **2005**, *30*, 65–72. [CrossRef]
- Scouller, K. The Influence of Assessment Method on Students' Learning Approaches: Multiple Choice Question Examination versus Assignment Essay. *High. Educ.* **1998**, *35*, 453–472. [CrossRef]
- Simkin, M.G.; Kuechler, W.L. Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decis. Sci. J. Innov. Educ.* **2005**, *3*, 73–98. [CrossRef]
- Iz, H.B.; Fok, H.S. Use of Bloom's Taxonomic Complexity in Online Multiple Choice Tests in Geomatics Education. *Surv. Rev.* **2007**, *39*, 226–237. [CrossRef]
- Harper, R. Multiple-Choice Questions—A Reprieve. *Biosci. Educ.* **2003**, *2*, 1–6. [CrossRef]
- Palmer, E.J.; Devitt, P.G. Assessment of Higher Order Cognitive Skills in Undergraduate Education: Modified Essay or Multiple Choice Questions? *BMC Med. Educ.* **2007**, *7*, 49. [CrossRef]
- Carpenter, S.K.; Pashler, H.; Vul, E. What Types of Learning Are Enhanced by a Cued Recall Test? *Psychon. Bull. Rev.* **2006**, *13*, 826–830. [CrossRef]
- Roediger, H.L.; Karpicke, J.D. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspect. Psychol. Sci.* **2006**, *1*, 181–210. [CrossRef]
- Greving, S.; Lenhard, W.; Richter, T. The Testing Effect in University Teaching: Using Multiple-Choice Testing to Promote Retention of Highly Retrievable Information. *Teach. Psychol.* **2022**, *0*, 1–10. [CrossRef]
- Larsen, D.; Butler, A.C. Test-Enhanced Learning. In *Oxford Textbook of Medical Education*; Walsh, K., Ed.; Oxford University Press: Oxford, UK, 2013; pp. 443–452.
- Wiklund-Hörnqvist, C.; Jonsson, B.; Nyberg, L. Strengthening Concept Learning by Repeated Testing. *Scand. J. Psychol.* **2013**, *55*, 10–16. [CrossRef]
- Karpicke, J.D.; Roediger, H.L. The Critical Importance of Retrieval for Learning. *Science* **2008**, *319*, 966–968. [CrossRef] [PubMed]
- Butler, A.C.; Roediger, H.L. Testing Improves Long-Term Retention in a Simulated Classroom Setting. *Eur. J. Cogn. Psychol.* **2007**, *19*, 514–527. [CrossRef]
- Butler, A.C.; Karpicke, J.D.; Roediger, H.L. The Effect of Type and Timing of Feedback on Learning from Multiple-Choice Tests. *J. Exp. Psychol. Appl.* **2007**, *13*, 273–281. [CrossRef] [PubMed]
- McDaniel, M. a.; Agarwal, P.K.; Huelser, B.J.; McDermott, K.B.; Roediger, H.L. Test-Enhanced Learning in a Middle School Science Classroom: The Effects of Quiz Frequency and Placement. *J. Educ. Psychol.* **2011**, *103*, 399–414. [CrossRef]
- Glass, A.L.; Ingate, M.; Sinha, N. The Effect of a Final Exam on Long-Term Retention. *J. Gen. Psychol.* **2013**, *140*, 224–241. [CrossRef]
- Pashler, H.; Bain, P.M.; Bottge, B.A.; Graesser, A.; Koedinger, K.; McDaniel, M.; Metcalfe, J. *Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007–2004*; National Center for Education Research, Institute of Education Sciences, U.S. Department of Education: Washington, DC, USA, 2007.
- Quizlet, L.L.C. Simple Free Learning Tools for Students and Teachers | Quizlet. 2022. Available online: <http://quizlet.com/> (accessed on 16 February 2022).
- Quitch. Quitch—Learning at your fingertips. 2022. Available online: <https://www.quitch.com/> (accessed on 12 September 2018).
- KEEUNIT. Quiz-App für Unternehmen. 2022. Available online: <https://www.keeunit.de/quiz-app-business/> (accessed on 28 February 2022).
- Klopfer, E. *Augmented Learning: Research and Design of Mobile Educational Games*; The MIT Press: Cambridge, MA, USA, 2008.
- Celador Productions Ltd. Who Wants to Be a Millionaire. 1998. Available online: <http://www.imdb.com/title/tt0166064/> (accessed on 14 December 2014).
- Bellis, M. The History of Trivial Pursuit. 2020. Available online: [http://inventors.about.com/library/inventors/bl\\_trivia\\_pursuit.htm](http://inventors.about.com/library/inventors/bl_trivia_pursuit.htm) (accessed on 20 April 2022).
- FEO Media, AB. QuizClash. 2022. Available online: <http://www.quizclash-game.com/> (accessed on 20 April 2022).
- Plain Vanilla. QuizUp—Connecting People through Shared Interests. 2013. Available online: <https://www.quizup.com/> (accessed on 12 January 2016).
- Russolillo, S. QuizUp: The Next “It” Game App? 2014. Available online: <http://www.palmbeachpost.com/videos/news/is-quizup-the-next-it-game-app/vCYDgf/> (accessed on 18 January 2016).
- Malone, T.W.; Lepper, M.R. Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning. *Aptit. Learn. Instr.* **1987**, *3*, 223–253.
- Garris, R.; Ahlers, R.; Driskell, J.E. Games, Motivation, and Learning: A Research and Practice Model. *Simul. Gaming* **2002**, *33*, 441–467. [CrossRef]
- Söbke, H. Space for Seriousness? Player Behavior and Motivation in Quiz Apps. In *Entertainment Computing, Proceedings of the ICEC 2015 14th International Conference, ICEC 2015, Trondheim, Norway, 29 September–2 October 2015*; Chorianopoulos, K.,

- Chorianopoulos, K.; Divitini, M.; Baalsrud Hauge, J.; Jaccheri, L.; Malaka, R., Eds.; Springer: Cham, Switzerland, 2015; pp. 482–489. [CrossRef]
30. Feraco, T.; Casali, N.; Tortora, C.; Dal Bon, C.; Accarrino, D.; Meneghetti, C. Using Mobile Devices in Teaching Large University Classes: How Does It Affect Exam Success? *Front. Psychol.* **2020**, *11*, 1363. [CrossRef]
  31. Kahoot! AS. Kahoot! 2022. Available online: <https://getkahoot.com/> (accessed on 12 January 2016).
  32. Showbie Inc. Socrative. 2022. Available online: <http://www.socrative.com> (accessed on 20 April 2022).
  33. Technische Hochschule Mittelhessen. ARSnova. 2014. Available online: <https://arsnova.thm.de/> (accessed on 13 March 2019).
  34. Fullarton, C.M.; Hoeck, T.W.; Quibeldey-Cirkel, K. Arsnova.Click—A Game-Based Audience-Response System for Stem Courses. *EDULEARN17 Proc.* **2017**, *1*, 8107–8111. [CrossRef]
  35. Basuki, Y.; Hidayati, Y. Kahoot! Or Quizizz: The Students' Perspectives. In Proceedings of the ELLiC 2019: The 3rd English Language and Literature International Conference, ELLiC, Semarang, Indonesia, 27 April 2019; EAI: Nitra, Slovakia, 2019; p. 202. [CrossRef]
  36. Wang, A.I.; Lieberoth, A. The Effect of Points and Audio on Concentration, Engagement, Enjoyment, Learning, Motivation, and Classroom Dynamics Using Kahoot! In Proceedings of the 10th European Conference on Game Based Learning (ECGBL), Paisley, Scotland, 6–7 October 2016; Academic Conferences International Limited: Reading, UK, 2016; p. 738.
  37. Wang, A.I.; Tahir, R. The Effect of Using Kahoot! For Learning—A Literature Review. *Comput. Educ.* **2020**, *149*, 103818. [CrossRef]
  38. Licorish, S.A.; Owen, H.E.; Daniel, B.; George, J.L. Students' Perception of Kahoot!'s Influence on Teaching and Learning. *Res. Pract. Technol. Enhanc. Learn.* **2018**, *13*, 9. [CrossRef]
  39. Christianson, A.M. Using Socrative Online Polls for Active Learning in the Remote Classroom. *J. Chem. Educ.* **2020**, *97*, 2701–2705. [CrossRef]
  40. Mendez, D.; Slisko, J. Software Socrative and Smartphones as Tools for Implementation of Basic Processes of Active Physics Learning in Classroom: An Initial Feasibility Study with Prospective Teachers. *Eur. J. Phys. Educ.* **2013**, *4*, 17–24.
  41. Kaya, A.; Balta, N. Taking Advantages of Technologies: Using the Socrative in English Language Teaching Classes. *Int. J. Soc. Sci. Educ. Stud.* **2016**, *2*, 4–12.
  42. Pechenkina, E.; Laurence, D.; Oates, G.; Eldridge, D.; Hunter, D. Using a Gamified Mobile App to Increase Student Engagement, Retention and Academic Achievement. *Int. J. Educ. Technol. High. Educ.* **2017**, *14*, 31. [CrossRef]
  43. Beatson, N.; Gabriel, C.A.; Howell, A.; Scott, S.; van der Meer, J.; Wood, L.C. Just Opt in: How Choosing to Engage with Technology Impacts Business Students' Academic Performance. *J. Acc. Educ.* **2019**, *50*, 100641. [CrossRef]
  44. IT Finanzmagazin. Quiz-App—Wissenszuwachs Durch Gamification: Wüstenrot Qualifiziert 1.400 Außendienstler. Available online: <https://www.it-finanzmagazin.de/quiz-app-wissenszuwachs-durch-gamification-wuestenrot-will-1-400-aussendienstler-qualifizieren-33706/> (accessed on 14 July 2016).
  45. Woods, B. QuizUp Launches Tools for Creating your Own Trivia Categories and Questions. 2015. Available online: <http://thenextweb.com/apps/2015/09/24/quizup-launches-tools-for-creating-your-own-trivia-categories-and-questions/#gref> (accessed on 12 January 2016).
  46. Kurdi, G.; Parsia, B.; Sattler, U. An Experimental Evaluation of Automatically Generated Multiple Choice Questions from Ontologies. In *International Experiences and Directions Workshop on OWL*; Springer: Cham, Switzerland, 2017; pp. 24–39.
  47. Howe, J. The Rise of Crowdsourcing. *Wired Mag.* **2006**, *14*, 1–16.
  48. Harris, B.H.L.; Walsh, J.L.; Tayyaba, S.; Harris, D.A.; Wilson, D.J.; Smith, P.E. A Novel Student-Led Approach to Multiple-Choice Question Generation and Online Database Creation, With Targeted Clinician Input. *Teach. Learn. Med.* **2015**, *27*, 182–188. [CrossRef]
  49. Papert, S. Situating Constructionism. In *Constructionism*; Papert, S., Harel, I., Eds.; Ablex Publishing: Norwood, NJ, USA, 1991.
  50. Moodle.org. Moodle. 2022. Available online: <https://moodle.org> (accessed on 23 May 2018).
  51. Campbell, E. MoodleQuiz: Learner-Generated Quiz Questions as a Differentiated Learning Activity; University of Dublin: Dublin, Ireland, 2010.
  52. Jones, J.A. Scaffolding Self-Regulated Learning through Student-Generated Quizzes. *Act. Learn. High. Educ.* **2019**, *20*, 115–126. [CrossRef]
  53. Guilding, C.; Pye, R.E.; Butler, S.; Atkinson, M.; Field, E. Answering Questions in a Co-Created Formative Exam Question Bank Improves Summative Exam Performance, While Students Perceive Benefits from Answering, Authoring, and Peer Discussion: A Mixed Methods Analysis of PeerWise. *Pharmacol. Res. Perspect.* **2021**, *9*, e00833. [CrossRef]
  54. Lin, X.; Sun, Q.; Zhang, X. Using Learners' Self-Generated Quizzes in Online Courses. *Distance Educ.* **2021**, *42*, 391–409. [CrossRef]
  55. Kurtz, J.B.; Lourie, M.A.; Holman, E.E.; Grob, K.L.; Monrad, S.U. Creating Assessments as an Active Learning Strategy: What Are Students' Perceptions? A Mixed Methods Study. *Med. Educ. Online* **2019**, *24*, 1630239. [CrossRef] [PubMed]
  56. Yu, F.Y.; Wu, W.S.; Huang, H.C. Promoting Middle School Students' Learning Motivation and Academic Emotions via Student-Created Feedback for Online Student-Created Multiple-Choice Questions. *Asia-Pacific Educ. Res.* **2018**, *27*, 395–408. [CrossRef]
  57. Teplitski, M.; Irani, T.; Krediet, C.J.; Di Cesare, M.; Marvasi, M. Student-Generated Pre-Exam Questions Is an Effective Tool for Participatory Learning: A Case Study from Ecology of Waterborne Pathogens Course. *J. Food Sci. Educ.* **2018**, *17*, 76–84. [CrossRef]
  58. Hancock, D.; Hare, N.; Denny, P.; Denyer, G. Improving Large Class Performance and Engagement through Student-Generated Question Banks. *Biochem. Mol. Biol. Educ.* **2018**, *46*, 306–317. [CrossRef]

59. Bottomley, S.; Denny, P. A Participatory Learning Approach to Biochemistry Using Student Authored and Evaluated Multiple-Choice Questions. *Biochem. Mol. Biol. Educ.* **2011**, *39*, 352–361. [\[CrossRef\]](#)
60. McClean, S. Implementing PeerWise to Engage Students in Collaborative Learning. *Perspect. Pedagog. Pract.* **2015**, *6*, 89–96.
61. Grainger, R.; Dai, W.; Osborne, E.; Kenwright, D. Medical Students Create Multiple-Choice Questions for Learning in Pathology Education: A Pilot Study. *BMC Med. Educ.* **2018**, *18*, 201. [\[CrossRef\]](#)
62. Yu, F.Y.; Liu, Y.H. Creating a Psychologically Safe Online Space for a Student-Generated Questions Learning Activity via Different Identity Revelation Modes. *Br. J. Educ. Technol.* **2009**, *40*, 1109–1123. [\[CrossRef\]](#)
63. Riggs, C.D.; Kang, S.; Rennie, O. Positive Impact of Multiple-Choice Question Authoring and Regular Quiz Participation on Student Learning. *CBE Life Sci. Educ.* **2020**, *19*, ar16. [\[CrossRef\]](#)
64. Khosravi, H.; Kitto, K.; Williams, J.J. RiPPL: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities. *J. Learn. Anal.* **2019**, *6*, 91–105. [\[CrossRef\]](#)
65. Wang, X.; Talluri, S.T.; Rose, C.; Koedinger, K. UpGrade: Sourcing Student Open-Ended Solutions to Create Scalable Learning Opportunities. In Proceedings of the L@S'19: The Sixth (2019) ACM Conference on Learning @ Scale, New York, NY, USA, 24–25 June 2019; Association for Computing Machinery: New York, NY, USA; Volume 17, pp. 1–10. [\[CrossRef\]](#)
66. Parker, R.; Manuguerra, M.; Schaefer, B. The Reading Game—Encouraging Learners to Become Question- Makers Rather than Question-Takers by Getting Feedback, Making Friends and Having Fun. In Proceedings of the 30th ascilite Conference 2013, Sydney, Australia, 1–4 December 2013; Carter, H., Gosper, M., Hedberg, J., Eds.; Australasian Society for Computers in Learning in Tertiary Education: Sydney, Australia, 2013; pp. 681–684.
67. Haladyna, T.M.; Rodriguez, M.C. *Developing and Validating Test Items*; Routledge: New York, NY, USA, 2013.
68. Ch, D.R.; Saha, S.K. Automatic Multiple Choice Question Generation from Text: A Survey. *IEEE Trans. Learn. Technol.* **2018**, *13*, 14–25. [\[CrossRef\]](#)
69. Söbke, H. A Case Study of Deep Gamification in Higher Engineering Education. In *Games and Learning Alliance, Proceedings of the 7th International Conference, GALA 2018, Palermo, Italy, December 5–7, 2018*; Gentile, M., Allegra, M., Söbke, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 375–386. [\[CrossRef\]](#)
70. Huang, B.; Hew, K.F. Do Points, Badges and Leaderboard Increase Learning and Activity: A Quasi-Experiment on the Effects of Gamification. In Proceedings of the 23rd International Conference on Computers in Education, Hangzhou, China, 30 November–4 December 2015; pp. 275–280.
71. Söbke, H.; Chan, E.; von Buttlar, R.; Große-Wortmann, J.; Londong, J. Cat King's Metamorphosis—The Reuse of an Educational Game in a Further Technical Domain. In *Games for Training, Education, Health and Sports*; Göbel, S., Wiemeyer, J., Eds.; Springer International Publishing: Darmstadt, Germany, 2014; Volume 8395, pp. 12–22. [\[CrossRef\]](#)
72. Chi, M.T.H.; Wylie, R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* **2014**, *49*, 219–243. [\[CrossRef\]](#)
73. Theobald, E.J.; Hill, M.J.; Tran, E.; Agrawal, S.; Nicole Arroyo, E.; Behling, S.; Chambwe, N.; Cintrón, D.L.; Cooper, J.D.; Dunster, G.; et al. Active Learning Narrows Achievement Gaps for Underrepresented Students in Undergraduate Science, Technology, Engineering, and Math. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 6476–6483. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Bates, S.P.; Galloway, R.K.; Riise, J.; Homer, D. Assessing the Quality of a Student-Generated Question Repository. *Phys. Rev. Spec. Top.-Phys. Educ. Res.* **2014**, *10*, 020105. [\[CrossRef\]](#)
75. Anderson, L.W.; Krathwohl, D.R.; Airasian, P.W.; Cruikshank, K.A.; Mayer, R.E.; Pintrich, P.R.; Raths, J.; Wittrock, M.C. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*; Pearson: New York, NY, USA, 2000.
76. Ni, L.; Bao, Q.; Li, X.; Qi, Q.; Denny, P.; Warren, J.; Witbrock, M.; Liu, J. DeepQR: Neural-Based Quality Ratings for Learnersourced Multiple-Choice Questions. *arXiv* **2021**, arXiv:2111.10058.
77. Weitze, L.; Söbke, H. Quizzing to Become an Engineer—A Commercial Quiz App in Higher Education. In Proceedings of the New Perspectives in Scienze Education, 5th Conference Edition, Florence, Italy, 17–18 March 2016; Pixel, Ed.; Webster srl: Padova, Italy, 2016; pp. 225–230.
78. Geiger, M.A.; Middleton, M.M.; Tahseen, M. Assessing the Benefit of Student Self-Generated Multiple-Choice Questions on Examination Performance. *Issues Account. Educ.* **2020**, *36*, 1–20. [\[CrossRef\]](#)
79. Caspari-Sadeghi, S.; Forster-Heinlein, B.; Maegdefrau, J.; Bachl, L. Student-Generated Questions: Developing Mathematical Competence through Online-Assessment. *Int. J. Scholarsh. Teach. Learn.* **2021**, *15*, 8. [\[CrossRef\]](#)
80. IJsselsteijn, W.A.; De Kort, Y.A.W.; Poels, K. The Game Experience Questionnaire. Technische Universiteit Eindhoven. 2013. Available online: [https://pure.tue.nl/ws/portalfiles/portal/21666907/Game\\_Experience\\_Questionnaire\\_English.pdf](https://pure.tue.nl/ws/portalfiles/portal/21666907/Game_Experience_Questionnaire_English.pdf) (accessed on 20 April 2022).
81. Voloshina, A. What Happened to QuizUp? 2021. Available online: <https://triviabliss.com/what-happened-to-quizup/> (accessed on 25 February 2022).
82. The University of Auckland | New Zealand. PeerWise. 2022. Available online: <https://peerwise.cs.auckland.ac.nz/> (accessed on 25 January 2022).