Students generate items for an online formative assessment: Is it motivating?

Rianne Poota, Renske A. M. de Kleijnb, Harold V. M. van Rijenc and Jan van Tartwijkd

aCenter for Teaching and Learning, Utrecht University, Utrecht, The Netherlands; bBiomedical Sciences, Education Center, University Medical Center Utrecht, Utrecht, The Netherlands

ABSTRACT
Background: A reported problem with e-learning is sustaining students’ motivation. We propose a framework explaining to what extent an e-learning task is motivating. This framework includes students’ perceived Value of the task, Competence in executing the task, Autonomy over how to carry out the task, and Relatedness.

Methods: To test this framework, students generated items in an online environment and answered questions developed by their fellow students. Motivation was measured by analyzing engagement with the task, with an open-ended questionnaire about engagement, and with the motivated strategies for learning questionnaire (MSLQ).

Results: Students developed 59 questions and answered 1776 times on the questions. Differences between students who did or did not engage in the task are explained by the degree of self-regulation, time management, and effort regulation students report. There was a significant relationship between student engagement and achievement after controlling for previous academic achievement.

Conclusions: This study proposes a way of explaining the motivational value of an e-learning task by looking at students’ perceived competence, autonomy, value of the task, and relatedness. Student-generated items are considered of high task value, and help to perceive relatedness between students. With the right instruction, students feel competent to engage in the task.

Introduction

E-learning

In the last few decades, e-learning has gained popularity in medical education (Ward et al. 2001; Ruiz et al. 2006). E-learning provides teachers with the opportunity to facilitate learning opportunities outside the classroom (Alonso et al. 2005). The effectiveness of e-learning has been thoroughly reviewed by Means et al. (2013), but teachers are experiencing difficulties in sustaining student-motivation when the students are preparing online for in-class activities (Keller & Suzuki 2004). Ideally students prepare for in-class activities without the incentive of their work being assessed. Reasons for avoiding assessing students are not only that it requires substantial effort of the teacher, but it also might have a negative impact on students’ intrinsic motivation (Deci & Ryan 2000).

Motivation

Students can either be more intrinsically motivated and, therefore, engage in an activity because they want to learn or to explore, or they can be more extrinsically motivated when their aim is to receive some kind of reward (Ryan & Deci 2000). Two influential motivational theories provide a framework for understanding why students are more intrinsically or extrinsically motivated. These are the expectancy-value theory (Wigfield 1994; Eccles & Wigfield 2002) and Ryan and Deci’s (2000) self-determination theory. The expectancy-value theory states that an individual’s intrinsic choice, persistence, and performance can be explained with two mechanisms: Students’ beliefs about how well they will perform on the activity and how much they value the task. So students’ intrinsic motivation and the way they are learning for a course can decrease when they think they will fail or do not think the task is valuable for them. According to the self-determination theory, the degree of intrinsic motivation is determined by the fulfillment of three basic psychological needs: Autonomy, Competence, and Relatedness. Ryan and Deci (2000) argue that when students do not feel that they are in control of their own learning processes (autonomy), do not think that they can execute the task (competence) or when they do not feel
respected, and cared for by others (relatedness), intrinsic motivation will decrease.

We propose a framework in which these two motivational theories are combined that teachers can use to estimate if an online activity will be considered motivating by students. This framework comprised four mechanisms: students’ perceived Value of the task, students’ perceived Competence over the task, and student’s perceived Autonomy over the task, and student’s perceived Relatedness regarding the task. In this article, we explore this framework with an e-learning activity that seems to meet the criteria of those four mechanisms.

**Student-generated items**

Teachers have experimented with the task of asking students to formulate multiple choice questions (MCQ) with the aim to stimulate formative self-assessment by the students (e.g. Fellenz 2004; Kolluru 2012; Papinczak et al. 2012). In the studies describing these experiments, students are generally asked to first, read the literature, second, formulate MCQs including the correct answer and multiple distractors, third, train themselves using their peers’ questions, and finally, provide feedback on the questions of their peers. Multiple studies have shown this activity of students generating items is an effective way of improving student achievement (e.g. Kerkman et al. 1994; Baerheim & Meland 2003). So far, in all the studies we found on this topic, the generation and answering of items was mandatory (e.g. Sircar 1999; Rhind & Pettigrew 2012). However, we hypothesize that student-generated items are also motivating when the task is not obligatory. Therefore, in this study, we explore if student-formulated items are motivating within a formative setting when the task-design is structured according to the four mechanisms. Students are instructed to only develop MCQs on the material that is obligatory for the assessment; this way of practicing with the questions has a high task value. Students receive proper instruction on how to design an adequate MCQ so that students will feel competent when doing the task. Students could choose whether or not to participate and when to participate so they perceive a high degree of autonomy. Lastly, students that participate are asked to give feedback on the questions of their peers in order to make the task more communal; as a result, students will perceive relatedness over the task.

**Aims and research questions**

Following motivational theories, we developed the hypothesis that students generating items in an online learning environment is a motivating task for students which results in active participation. Additionally, we are seeking confirmation that active participation results in higher academic achievement even when the task is not obligatory. Therefore, this study addresses the following research question: To what extent is the practice of student-generated questions in an online environment motivating for all students, and what are the effects on student outcomes?

The most important contributions of this study will be (a) insight in the motivational value of student-generated questions for different students and (b) a (dis)confirmation of the effectiveness of student-generated items on achievement.

**Methods**

This study was conducted within a ten-week, second year, mandatory undergraduate physiology course in a biomedical science program. The course comprised three parts, each focusing on a physiology-oriented topic: respiratory, circulatory, and urinary organ systems. Each part had a duration of three weeks and was assessed with a summative MCQ-test at the end of the third week. In the tenth week, an overall open summative assessment was given covering the entire course.

**Participants**

In total, 159 undergraduate biomedical students registered for the course. At the beginning of the second part of the course (week 4), students were introduced to student-generated questions and the online learning environment. Students were given a written statement about the aim of the research involving their participation in PeerWise, their motivation, their learning strategies and their learning outcomes, and were asked for their written informed consent. A total of 109 students (69%) participated in this study (37 men and 72 women, median age 20 years) and signed the “informed consent” form. There was no personal reward granted for partaking in the research.

**Data gathering**

**Motivated behavior**

In order to measure motivated behavior in the course, we first analyzed all online activity of the students regarding constructing and practicing MCQs in the online environment. To track all activity, we used PeerWise, an online learning environment that builds an online database with all contributed test-items (Denny et al. 2008). Second, we asked students why they showed certain behavior. Students were also asked to answer the following open questions regarding their activity in the online learning environment: “Why did you or did you not make an account in PeerWise?” and “Why did you or did you not develop MCQs?”

**Student’s self-reported motivation and learning strategies**

Self-reported motivation was measured with the six motivation scales of the motivated strategies for learning questionnaire (MSLQ: Pintrich et al. 1991): Intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, self-efficacy for learning and achievement, and test anxiety. Items of the MSLQ were formulated as statements that had to be rated on a five-point Likert-type scale, ranging from “1 = Does not apply to me at all” to “5 = Fully applies to me.” A translated and validated Dutch version of the MSLQ was used (Bouwmeester et al. 2016).

Learning strategies were measured with the nine scales of the MSLQ: Rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time and study
environment management, effort regulation, peer learning, and help seeking. In Table 1, a sample question for all 15 subscales is presented with the Cronbach’s alpha found by Pintrich et al. (1991) and the Cronbach’s alpha found for this study. Cronbach’s alpha in this study ranged between 0.57 and 0.86. All scales for which the Cronbach’s alpha was lower than 0.70 were not further taken into account due to low reliability (Nunally 1979).

### (Previous) academic achievement

To measure students’ achievement after the intervention, students’ grades on the second summative test in the course were used. In order to use previous achievement as a covariate, students’ grades on the first summative test in the course were used.

### Procedure

In the first part of the course, students were given a teacher-constructed formative test to prepare for the first summative test. At the beginning of the second part of the course, the students were given a presentation about all requirements and the instructor’s teaching style. Additionally, two rules for participation were provided on how to develop higher cognitive MCQ (cf. Bloom et al. 1956). Additionally, two rules for participation were given. First, constructed questions have one correct answer and three distractors and include an explanation for why the possibilities given are either correct or wrong. Second, after practicing a question, a rating of the question’s difficulty and quality should be given. To provide a sense of autonomy, students were free to decide if they wanted to participate. However, in order to stimulate participation, the faculty vowed that if students would contribute more than 150 unique questions in PeerWise, five of those questions would be included in the summative assessment, which consisted of 40 items. Students were free to develop questions on the 14 core topics of the course preceded by the faculty, and they could practice with the questions constructed by their peers for the full three weeks before their small summative test.

After the final open test in week 10, students were asked to answer the open questions and the MSLQ on paper.

### Analysis

The tracking data, the data gathered with the MSLQ, and the results on the four summative tests of 109 students were inserted in one database. Students were categorized into two main groups using the tracking data: Students that showed motivational behavior and thus did not participate. Within the group that showed motivational behavior, a distinction was made between students that both developed and answered questions and students that only answered questions.

First, the tracking data were analyzed by calculating the mean and standard deviations for the developed questions and the answered questions. Furthermore, answers to the two open questions were analyzed based on the theoretical framework of motivation.

Second, a multiple ANOVA was used to test for differences in the five motivation scales and eight learning strategies scales between the two main groups. Since, the size of the groups was different, a Gabriel post-hoc test was used.

Third, an ANCOVA was used to test for differences in achievement on the second summative test between the

### Table 1. Sample questions for subscales MSLQ.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>n</th>
<th>Sample question</th>
<th>α Pintrich et al. (1993)</th>
<th>α This study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>4</td>
<td>In a class like this, I prefer course material that really challenges me so I can learn new things.</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>Extrinsic motivation</td>
<td>4</td>
<td>If I can, I want to get better grades in this class than most of the other students.</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>Task value</td>
<td>5</td>
<td>I think the course material in this class is useful for me to learn.</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>Control</td>
<td>4</td>
<td>If I do not understand the course material, it is because I did not try hard enough.</td>
<td>0.68</td>
<td>0.60</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>7</td>
<td>I’m confident I can do an excellent job on the assignments and tests in this course.</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>Test anxiety</td>
<td>5</td>
<td>When I take tests I think of the consequences of failing.</td>
<td>0.80</td>
<td>0.84</td>
</tr>
<tr>
<td>Learning strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rehearsal</td>
<td>4</td>
<td>I make a list of important terms for this course and memorize the lists.</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>Elaboration</td>
<td>5</td>
<td>When reading for this class, I try to relate the material to what I already know.</td>
<td>0.75</td>
<td>0.60</td>
</tr>
<tr>
<td>Organization</td>
<td>4</td>
<td>I make simple charts, diagrams or tables to help me organize course material.</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td>Critical thinking</td>
<td>5</td>
<td>I treat the course material as a starting point and try to develop my own ideas about it.</td>
<td>0.80</td>
<td>0.71</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>11</td>
<td>I try to change the way I study in order to fit the course requirements and the instructor’s teaching style.</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>Time management</td>
<td>8</td>
<td>I make good use of my study time for this course.</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>Effort regulation</td>
<td>4</td>
<td>I work hard to do well in this class even if I do not like what we are doing.</td>
<td>0.69</td>
<td>0.73</td>
</tr>
<tr>
<td>Peer learning</td>
<td>3</td>
<td>I try to work with other students from this class to complete the course assignments.</td>
<td>0.76</td>
<td>0.68</td>
</tr>
<tr>
<td>Help seeking</td>
<td>3</td>
<td>I ask the instructor to clarify concepts I do not understand well.</td>
<td>0.52</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Bold values indicate a Cronbach’s alpha ≥ 0.70.
two main groups. As a covariate, students’ achievement on the first summative test was used.

The student-generated questions were exported from PeerWise, and each item was labeled to be a lower order-thinking question or a higher order thinking question (cf. Anderson & Krathwohl 2001).

Results

Of the 109 students that participated in this study, 45 students showed motivational behavior: 22 of those 45 students developed questions ($M = 2.68$, $SD = 2.70$), and 41 of them answered questions ($M = 43.32$, $SD = 33.74$). In total, 59 questions were developed by the 22 students who developed questions. The 41 students answered 1776 questions. Of the 59 developed questions, 11 were labeled as higher order thinking questions (cf. Bloom et al. 1956). About one-third of the lower order thinking questions were pure factual recall (often in the format of a true–false question), but the bulk of the lower order questions regarded the understanding level of Bloom. In Figure 1, a few examples of higher order questions and lower order questions are represented.

The information that was gathered with the open questions can be summarized as follows:

- Of the group of students that did not show motivational behavior ($n = 59$), nearly 60% had the intention to participate but wrote that in the end, they did not participate due to a lack of time (“I really wanted to do the assignment because it could help me, but in the end, I did not find the time”). Other reasons were that they had forgotten about the task, did not see the value of the task or did not feel competent enough to develop questions.
- Practically all students that did show motivational behavior ($n = 45$) wrote that they wanted to answer the questions in order to increase their competence (“I wanted to check if I understood all questions from my peers to see if I was ready for the test”).
- Some of the students that developed questions ($n = 22$) also wrote that they thought this would improve their competence. However, most of the students felt strongly that before they could practice the questions of peers, they also had to develop some questions themselves that could help others. (“I strongly believe that if you want something from your peers, you also have to give something in return”).

A one-way ANOVA between groups was conducted to compare self-reported motivation and learning strategies of students that showed motivation and did not show motivation. A main effect was found for the learning strategies of self-regulation ($F(1105) = 4.87$, $p = 0.030$), time management ($F(1105) = 21.66$, $p < 0.001$), effort regulation ($F(1100) = 18.82$, $p < 0.001$), and help-seeking ($F(1105) = 4.45$, $p = 0.037$). So, unlike our hypothesis, we did not find a difference in motivational attitudes between students that participated and did not participate. However, it was found that students that participated scored significantly higher on four learning strategies from students that did not participate: Self-regulation, time management, effort regulation, and help seeking.

An ANCOVA was used to test for differences in achievement on the second summative assessment between the two groups. As a covariate, the achievement of the first summative assessment was used. There was a significant difference on achievement between the motivated ($M = 7.2$) and non-motivated ($M = 6.0$) groups, even after controlling for previous academic achievement ($F(1106) = 11.48$, $p = 0.001$). These findings were consistent with the hypothesis that when students were participating in developing and answering student-generated items, they
would have a higher achievement on the summative exam, even when accounting for previous knowledge.

Discussion

Our findings suggest that, contrary to our expectations, student-generated items following the design in which the four mechanisms of task-value, autonomy, competence, and relatedness were incorporated cannot be considered motivating for all students. In order to understand this finding, we will take a more specific look into all four mechanisms.

**Autonomy:** We studied to what extent students showed motivational behavior after letting students decide if they wanted to engage in the task and what explanations they provided for this (lack of) motivational behavior. Slightly, over half the students did not show motivational behavior.

**Task value:** When asked why students did not participate, about 60% answered that, they did have the intention to do so since they saw the value of the task, but they ultimately did not do so due to lack of time. The students that did perform the task by answering student-generated items explained that the value of the task was the most important reason to participate.

**Relatedness:** For the vast majority of students that both answered and developed questions, the main reason for participation was the need to do something for their peers.

**Competence:** We may conclude that students felt sufficiently competent to perform the task since only three students stated that they felt they were not ready to develop good questions.

So, reasoned from the answers provided by the students, student-generated items have a high task value, and students feel competent in performing the task and contribute more because they felt that they had to help their peers forward. However, maybe due to too much autonomy, not all students who were motivated to participate finally engaged in the task.

To understand why half of the students did not participate in formulating or answering student-generated questions, we looked for differences between the groups on motivation and learning strategies. We concluded that the two groups indeed did not differ in their motivations for the course; however, they did differ on three learning strategies: Self-regulation (the ability to plan, monitor, and regulate a task), time management (the ability to manage time effectively), and effort regulation (the ability to control one’s efforts and committing to one’s own set goals) (Pintrich et al. 1993). This confirms our findings reported above when looking at the motivational behavior and students’ explanations. Specifically, the group of students that actively executed the task found time and energy to adhere to their original plan and, therefore, can be described as more conscientious students (Poropat 2009). So with this design, several students do not do the task because of lack of regulatory skills, not because they did not think the task was motivating.

Lastly, we studied if students that engage in student-generated questions achieve better than their peers who did not engage. Confirming earlier studies concerning student-generated items (Palmer & Devitt 2006; Papinczak et al. 2012), we found that students that engaged in the task performed higher on the summative assessment, even when corrected for earlier academic achievement. This result means that developing and practicing student-generated questions leads to better academic achievement in both conditions where the task is mandatory (and so low in autonomy) and where the task is free to choose (and so high in autonomy).

With this study, we have gained insight in the motivational value of student-generated questions in which the value of the task and the need for relatedness are the two greatest motivational drives for students to partake. Also, we have found that students that participate in the task can be described as more conscientious than their peers, and we have confirmed that students that participate in the task do have higher academic achievement.

It is important to keep in mind two limitations of this study. Although the motivation segment of the MSLQ is comprised six scales, the reliability of the internal motivation scale was poor. Therefore, we could not take this scale into account when analyzing student motivation. Furthermore, there were large differences in the number of questions students developed and answered. It would have been interesting to investigate differences in motivation and learning strategies within the group of students that participated. However, this was not possible due to the small number of students who developed questions and thus due to lack of power.

Following the results and limitations of this study, two themes for future research could be further explored. The first theme considers the further investigation of the design in order to better understand the underlying mechanisms. Considering autonomy, it would be interesting to develop the student-generated questions in a mandatory situation. In this way, autonomy of students is reduced in order to provide less conscientious students (and, therefore, students with poorer time management skills) with better opportunities to also engage in the task. This may lead to more time-on-task for preparing in the summative exam and, therefore, better summative performance. Considering task value, it would be interesting to vary the proportion of student-generated questions in the final assessment in order to see if this is more motivating.

The second theme considers, the four mechanisms in this study that need to be present in order for students to find a task motivational: competence, autonomy, value of the task, and relatedness. This proposed framework should further be explored and tested in other e-learning situations in order to determine if the framework can be widely used to understand if a task is motivational.

Conclusions

This study proposes an effective way of investigating the motivational value of an e-learning task by looking at students’ perceived competence, autonomy, value of the task, and relatedness. Student-generated items are considered of high task value, help to perceive relatedness between students, and when given the right instruction, students feel competent to engage in the task. To provide less conscientious students a better opportunity to engage in the tasks, teachers could aid students by allocating time in the course to work on the task. In this way, all motivated
students can benefit from the task, which ultimately results in better academic achievement.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Glossary

**Student-generated items**: The task of asking students to formulate multiple choice questions (MCQ) with the aim to stimulate formative self-assessment by the students. Students are generally asked to first read the literature, secondly formulate MCQs including the correct answer and multiple distractors, thirdly train themselves using their peers’ questions, and finally, provide feedback on the questions of their peers. Multiple studies have shown this activity of students generating items is an effective way of improving student achievement.

Notes on contributors

*Rianne Poot*, MSc, is an Educational Researcher and Educational Consultant at the Center for Teaching and Learning, Utrecht University, The Netherlands.

*Renske A. M. De Kleijn*, PhD, is a Post-Doctoral Researcher and Educational Consultant at the Center for Teaching and Learning, Utrecht University, The Netherlands.

*Harold V. M. Van Rijen*, PhD, is a Professor of Innovation in Biomedical Education, University Medical Center Utrecht, The Netherlands.

*Jan W. F. Van Tartwijk*, PhD, is a Professor of Education, Faculty of Social and Behavioral Sciences, University Utrecht, The Netherlands.

**References**


