**Question Generation as a Strategy for
the Advancement of Intermediate- and Low-Achieving Students**

**ABSTRACT**

One of the most important goals in education is the nurturing of higher-order thinking among all students, of all ages and all levels. Student question generation (SQG) is a teaching and learning strategy that promotes higher-order cognitive skills. The purpose of this study is to examine which students gain from SQG activities. Is it mainly those with strong academic achievements, or do lower-achieving students profit as well? The study was performed over a six-year period during which 171 science teacher trainees in Israel generated, answered, and peer-evaluated questions at various orders of thinking. The students were divided into four groups commensurate with their level of achievement in question-generation before the experiment: very good, good, intermediate, and poor. When the pre- and post-SQG test scores were examined, it was found that the question-generation activities contributed the most to intermediate- and low-achieving participants, whose overall scores showed statistically significant improvement. Furthermore, an isolated comparison of test scores for answering higher-order-thinking questions (pre- and post-SQG) indicated an even greater improvement among intermediate- and low-achieving students in their ability to cope with such questions. The findings are important in changing commonly held attitudes among teachers about the ostensible inability of underachieving pupils to engage in higher-order thinking tasks. If such students are given encouragement, they may be enlisted in such tasks.

Keywords: student question-generation; low-achieving students; higher-order-thinking questions; intermediate-achieving students

**INTRODUCTION**

Student question generation (SQG) as well as the solving and evaluation of such questions are active tasks that promote various aspects of cognitive and metacognitive learning ( ). Despite their importance, however, teachers do not tend to include them in their work, instead focusing students’ attention on answering questions as opposed to generating them.

This study is the continuation of a previous study among science teacher trainees that proposed a model for the integration of SQG activities. The study also presented findings indicating that practicing SQG according to this model helped to improve the students’ ability to cope with higher-order-thinking questions.

A concise version of the model for integrating SQG activities, described at length in the previous study, is included as an appendix to this study. Generally speaking, the model integrates three different activities: (1) student question-generation (2) students answering peer-generated questions; and (3) peer assessment of other students’ questions.

The study described below took place over a six-year period. The research population was divided into four groups differentiated by their achievements before the SQG activity. The main research question was: Which groups of students benefited the most from SQG—was it mainly the high-achievers, or did lower-achieving students also gain?

Importantly, most studies about low-achieving students are performed among primary- or secondary-school pupils and not among college students, as is the case in the current study. The study is also important in that, while many studies have dealt with inquiry-based learning and problem-solving, few have addressed SQG; and among those dealing with SQG, very few examine its effect on pupils of varying levels of achievements.

**STUDENT QUESTION GENERATION**

Questioning is fundamental in learning, and students’ questions play a crucial role in meaningful learning and learning motivation.The types and levels of questions may be classified commensurate with the order of thinking that is needed to answer them. One of the most commonly accepted classifications is Bloom’s Taxonomy (Bloom et al., 1956), which offers a hierarchy for different kinds of questions ranging from knowledge questions, which reflect the lowest order of thinking, to comprehension questions, application, analysis, synthesis, and evaluation. Decades later, Anderson and Krathwohl (2001) revised the taxonomy by emphasizing differences among the cognitive processes. Other differentiations are broader in their reference, specifying a general classification of higher-order and lower-order questions. Papinczak et al. (2012), for example, sorted questions into two groups: confirmation questions and transformation questions. Confirmation questions are meant to elucidate information and define and explain concepts, whereas transformation questions involve reconstruction and reorganization of the student’s comprehension. Transformation questions are considered examples of higher-order-thinking questions. They include, for example, Bloom, Anderson, and Krathwohl’s application, analysis, synthesis, and evaluation questions.

The pedagogical value and importance of SQG is firmly based in empirical findings. A comprehensive analysis of 109 empirical studies on SQG, conducted in numerous disciplines and across all age groups (from primary school to college), yielded a widespread consensus on SQG’s positive effects on learning (Yu, 2012). An extensive study among science students at three different universities in Britain, for example, examined the effect of three student activities associated with multiple-choice questions: answering questions, generating questions, and examining and commenting on peers’ questions. A significant positive correlation was found between these activities and test scores when all three activities took place (Hardy et al., 2014). Similarly, a study of 10th grade science pupils found that those who practiced question-generation improved their questioning ability as well as their academic achievements. Moreover, the findings showed that question-generating skills can be used as an alternative assessment method, particularly in order to assess higher-order thinking (Dori and Herscovitz, 1999; Offerdahl and Montplaisir, 2014). Similarly, Koch and Eckstine (1991) found that college physics students improved their reading comprehension when they were taught question-generation skills. Learning this skill stimulated their self-awareness of difficulties in reading comprehension and may have served as self-regulated learning.

Research has found that students who invoke self-regulated learning processes refine their learning skills and develop critical thinking (Nguyen and Ikeda, 2015; Stefanou et al., 2013). Question generation is an important metacognitive strategy that focuses students’ attention on content and on central ideas, helping them develop critical thinking, self-critique, and creativity (Chin and Brown, 2002; Rothstein and Santana, 2011). Interestingly, one study compared the effect on students of engaging in answering questions versus engaging in generating questions, examining their academic achievements and cognitive and metacognitive strategies. This study found no differences in academic achievements between students who answered questions and those who generated them; both activities were found to be equally effective. Students who engaged in generating questions, however, exhibited significantly higher-level cognitive strategies and metacognitive skills than did the others. They were more aware of their learning process, were better able to critique their own work, and self-assess their progress, and more willing to change (Yu and Liu, 2008).

Although most studies demonstrate the value of SQG in the advancement of learning, this activity has hardly been incorporated into scholastic settings. Many lessons tend to be teacher-controlled monologues. Students in class ask few questions (Nystrand et al., 2003) and most questions that they ask are basic knowledge questions that entail regurgitation of information (Chin and Brown, 2002; Middlecamp and Nickel, 2005). Particularly in higher education, students focus on questions that teachers ask or that are harvested from textbooks. Self-generated questioning, foremost the kind that involves higher-order thinking, is a process that most students practice only to a limited extent (Dori and Herscovitz, 1999; Yu and Chen, 2014).

Several reasons for students’ limited question generation have been proposed. Teachers who do not feel confident enough in their discipline suppress questioning. Class atmosphere, pupils’ fear of answering incorrectly, and teacher–pupil relations all influence questioning (Dillon, 1988). The number and type of questions that students ask depends on numerous additional factors, such as students’ age, experience, and skills, the nature of and interest in the subject studied, and students’ proficiency in the subject (Shodell, 1995).

**LOW-ACHIEVING STUDENTS AND HIGHER-ORDER THINKING**

Nurturing higher-order thinking among students of all ages and at all levels is one of the most important goals in education (Yang, 2016; So, 2010). Many teachers, however, believe that tasks entailing higher-order thinking are suited mainly to high-achieving students and are beyond the reach of low achievers whose command of the basic information is weak (Raes, 2014; Zohar, 2001). This conventional wisdom among teachers is predicated on the idea that knowledge-building is hierarchical: Only after knowledge of the subject is mastered can students advance to tasks of higher cognitive orders of comprehension or application in the same subject. Due to this hierarchical conception of knowledge-building, low-achieving pupils are often assigned low-cognitive-level tasks only (Shepard, 1991).

However, this hierarchy has been blurred by developments in understanding how knowledge is built, with emphasis on comprehension and thinking at all levels of learning, including the acquisition of basic knowledge. It has been claimed that the traditional concept, by which thinking and understanding can take place only after the base is mastered, can no longer play a leading role in teaching and that the advancement of thinking must be applied in all learning and to all students (Bruer, 1992; Bransford, 2000).

Low-achieving students often have learning difficulties (Zohar, 2003), limited learning skills, and poor self-image (Hacker, 2000). One of the key questions concerns the extent to which learning strategies that entail high-level cognitive skills can contribute to these students. Motivating students on different levels to contend successfully with higher-order thinking when learning is certainly a challenge for teachers. Several studies, however, show that it is possible. White and Fredrickson (1998), for example, demonstrated that both high-achieving and low-achieving students gain from shared inquiry-based learning. They found that assessment by means of a structured portfolio had a positive effect on students’ comprehension and that this effect was even stronger among low-achievers.

Anat and Dori (2003) reinforced these findings by showing the favorable effect of co-learning on the performance of both high-achieving and low-achieving students. In one of their studies on high-school students, they even demonstrated that low-achievers profit more significantly from reflective inquiry than do their high-achieving peers. Similarly, Raes et al. (2014) found that co-learning online contributed to all pupils but did more for the advancement of low-achievers.

Studies among college students show that even at this stage of schooling, high-thinking-level tasks are effective and contribute to students at all levels. Chiu and Cheng (2017), looking into the effect of active learning classrooms on university students in Hong Kong, found that those with high, intermediate, and low achievements all gained from active learning. The students’ creativity levels improved irrespective of their academic achievements (Chiu, 2017). Kogan and Laursen (2014) showed that mathematics students at four colleges made more progress in courses that employed inquiry-based learning than in other courses, with the strongest effect of inquiry-based learning found among students with low achievements.

A few studies, mostly performed in primary and secondary schools, looked into the direct effect of SQG activities on students at different levels. Kaya (2015), for example, reported that high-achieving pupils at a primary school generated more questions, and on a higher order of thinking, than did low-achieving peers. Yerrick (2000) showed that underachieving high-school students who had lengthy histories of failures in school and were involved in generating questions and designing an experiment underwent meaningful changes in their comprehension of scientific-inquiry processes and the nature of science.

The current study is unique because it tests the effect of SQG activities on college students at various levels. In a previous study by the researcher, it was found that practice in generating, answering, and evaluating questions helps to improve the achievements of some college students. The current study, as a continuation of the previous one, asks who gained more from these activities—was it mainly high achievers, or did low achievers also profit?

**RESEARCH QUESTIONS**

1. Did practice in generating, answering, and assessing questions abet an improvement in overall test scores among each of the four groups of students—low, intermediate, good, and very good achievements—and, if so, to what extent?
2. Did practice in generating, answering, and assessing questions help to improve the scores on higher-order questions within each of the four groups of students—and if so, to what extent?

**METHODOLOGY**

**Research Approach**

The research is based on comparative testing for pre- / post-test intervention. The intervention pertains to students who engaged in generating, answering, and peer-assessing questions in a college cell-biology course.

**Participants**

The research population comprised nine classes of science-education students who took a cell-biology course at two academic education colleges in Israel. Both colleges are located in the south of Israel, are attended mainly by females, and have very similar science-education curricula. Class-size ranged from fifteen to twenty-seven, and a total of 171 students participated in the study (152 women, 19 men) with an average age of 21.9 years. All the students took the same course syllabus with the same lecturer, who had approximately twenty-two years of teaching experience.

**Procedure**

The study was conducted over a period of six academic years from 2010–2016. The cell-biology course was taught in each of the nine classes for two semesters, the first semester from October to January and the second from March to June. Lessons in each course were given once a week for two hours—a total of fifty-six hours per course, with fourteen lessons per semester. In the first semester, the students did not engage in question generation. In the last lesson of the first semester, time was set aside to study for the course exam; students were given examples of questions and an opportunity to ask questions about each topic studied during the semester. Throughout the second semester, the students engaged in question-generation activities as described in length in the previous study (a description of these activities is also attached to this study as an Appendix).

Exams were used in order to examine the effect of the SQG activities: At the end of each semester, the students were tested on the subject matter that was covered during the semester. Each exam comprised approximately fifteen questions, most of which (around eleven) were closed confirmation questions involving knowledge and memory; four questions (about 25 percent of the exam) were open transformation questions testing comprehension, application, or synthesis. All nine classes took exams that were very similar, with minor variations.

**Data Analysis**

After the first-semester exams were checked, the students in all of the classes were sorted into four achievement groups: low, intermediate, good, and very good. Table 1 itemizes the range, distribution, and frequency of each group’s scores. To confirm the aforementioned grouping of the students, the students’ scores from two additional first-semester courses were checked; both courses were in the Life Sciences (Zoology and General Botany). For 140 students—82 percent of the total—a fit was found. Namely, the three scores of each of these students fell into the same range, as specified in Table 1. For the other eighty-nine students, whose three scores fit into more than one range or one group, the three scores were averaged to determine the student’s placement in the appropriate group.

Subsequently, the first- and second-semester exams were scored and the averages and standard deviations of each group’s scores were calculated. The four higher-order-thinking questions were scored on a standard scale: points were given for the accuracy of the answers, a description of the explanation, and the reasoning. For each student and each group, the average score and standard deviations for the higher-order-thinking questions were calculated. About 20 percent of all exams (thirty-four exams, at least five from each group) were scored by an additional lecturer with extensive experience in cell biology. The correlation among the scores was a strong 87 percent.

To make sure that the students’ apportionment into four achievement-based groups lent itself to reference and statistical comparison, an ANOVA test to test inter-group variance was performed. Duncan post-hoc tests revealed significant differences among each of the four groups in regard to all independent variables (Table 2). By inference, the groups may be treated as distinct.

To compare the first-semester and second-semester scores of each student and each group, paired t-tests were performed, between the overall exam scores in each semester and between scores for the higher-order-thinking questions alone.

**FINDINGS**

**The Effect of SQG on Overall Scores for the Groups of Students**

Table 3 presents the comparison of each group’s first-semester exam scores—pre- SQG—and their second-semester scores, post-SQG. It can be seen that there was no statistically significant increase in exam scores among good-achieving and very-good-achieving students. However, a statistically significant upturn in post-SQG scores was evident among both low-achieving and intermediate-achieving students. The increase in the average score of the low-achieving group was particularly impressive—21 percent compared with 9 percent among the intermediate-achieving cohort.

*Place Table 3 here*

**The Effect of SQG on Scores for Higher-Order-Thinking Questions among the Student Groups**

A comparison of only the higher-order-question scores on the pre-SQG and post-SQG exams (Table 4) shows that most students’ scores on the thinking questions rose significantly after the activity. In fact, only the group of highest achieving students did not improve its score significantly. As in the overall score, the low-achieving student group recorded the strongest improvement.

*Place Table 4 here*

**DISCUSSION AND IMPLICATIONS**

In the previous study, it was found that activities in generating, answering, and peer-assessing questions are not conducive to significant improvement in the total score of an entire student population ( ). The current study, however, which segmented the students into groups that were differentiated by academic achievements, shows that specifically those with low and intermediate achievements improved their overall scores significantly. Furthermore, the analysis that compares students’ achievements on higher-order-thinking questions alone points to an especially significant and impressive improvement among students of intermediate and low achievements.

The overall findings of the current study indicate that most students gained from the SQG activity. These findings reinforce Yu (2012), who in a far-reaching overview analyzed numerous studies among students of different ages and different disciplines that reported a favorable effect of SQG. However, what was not examined in those studies and was found in the current study is that SQG activities have the greatest benefit to low- and intermediate-achieving students, improving their cognitive ability to contend with higher-order-thinking questions.

These encouraging findings reinforce the belief that all students should be encouraged to tackle higher-order tasks. Those with limited learning skills and poor self-image, however, need more systematic and structured support (Zohar, 2003). Absent such support, the strong correlation between poor learning capabilities and low scores may persist (Proctor, 2006). Such support may take the form of active co-learning. In studies on the effect of shared inquiry-based learning, for example, the strongest effect was found among low-achievers (Kogan and Laursen, 2014; Raes, 2014). Generating and assessing questions—particularly higher-order-thinking questions—entails complex thinking skills, much as inquiry-based learning does. Furthermore, the SQG activities were conducted in groups as shown; even relatively brief practice with these activities had a dramatic effect on the students who had the greatest difficulties.

Thus, weak students in the current study obtained support chiefly through active learning and co-learning. Shared generation, answering, and assessment of questions, and sharing of knowledge by means of a question bank were especially useful anchors for the weakest students. The question bank that was amassed from the full set of student-generated questions was instrumental in studying for exams, alleviating anxiety before exams, and made similar test questions (author) [כן? "שלי"?] easier to cope with. Interaction and co-learning confer valuable cognitive and metacognitive advantages. Question-generation forces one to attain strong command of the material and peer assessment evokes reflection about one’s own learning and lends it greater depth (Hsiung, 2012; Chin and Osborne 2008).

Importantly, support for weak students should not be overly demanding; lecturers should not be burdened with mobilizing all of the students to contend with complex tasks. College lecturers may refrain from active teaching strategies that entail extensive preparation or the sort that do not leave enough time to teach everything in the syllabus. SQG activities encourage students to play an active role in learning without imposing a special burden on the lecturer and necessitating significant changes in course topics. This is what gives them strong potential for assimilation and adoption. In addition, SQG activities are appropriate for college students who focus on the goal of doing well on examinations. The creation of a question bank with which to prepare for an exam is the most important factor in mobilizing students to learn by generating, answering, and peer-assessing questions (author).

The contribution of SQG to the low-achieving group is especially important because these students are, by and large, the most prone to quitting their studies. An improvement in their scores often means crossing from “fail” to “pass” in a given course. This would reinforce their self-confidence, their belief in their ability to succeed, and their motivation—abetting the retention of students who might otherwise drop out (Respondek et al., 2019).

Most participants in the current study were students in their first-year of an academic degree—a stage when support is particularly important. First-year students must adjust to unfamiliar learning environments, cope with challenging tasks, and surmount stresses occasioned by failure to satisfy requirements (Perry, Hall, et al., 2005). After experiencing failure, most first-year students report a decrease in their ability to continue being successful in studies, in contrast to what they report after performing tasks successfully (Hall, 2008). The student dropout rate changes with each year of studies, with the highest rate found in the first year (Respondek et al. 2019; Alarcon and Edwards, 2013). Therefore, a meaningful improvement in scores after SQG activities may enhance the weakest students’ confidence in scholastic success and make it less likely that they will drop out.

In this study, the highest-achieving students improved neither their overall scores nor their scores on higher-order questions after SQG. This is unsurprising because this group had less room for improvement to begin with. Nevertheless, it is important to challenge these students, too, with complex thinking tasks and to motivate them to contribute to the group’s learning.

One of the limitations of this study is that the comparison between the first-semester (pre-SQG) score and the second-semester (post-SQG) score involved exams that covered different subject matter, as taught in each semester. Different subject matter, even in the same discipline, may affect the student’s degree of comprehension and ability to cope with exam questions. The alternative, however—comparing different students studying the same subject matter—would have created an even more significant research limitation due to student variance. Another limitation was the relatively short duration of the SQG experience. A lengthier and more intensive activity might have yielded a broader improvement in scores.

In sum, mobilizing students on different levels to cope successfully with learning at a higher-order of thinking is a cardinal issue in research on teaching and learning. SQG activity is a constructive learning and teaching strategy that abets the advancement of low- and intermediate-achieving students in particular. These findings are important in changing the widely held belief among teachers that underachieving students cannot handle higher-order tasks—a belief that thwarts these students’ progress and adversely affects the narrowing of disparities and the assurance of equal opportunity. Practically speaking, the SQG model may be integrated into almost every discipline taught at the college level and not only for science teacher trainees. The more such activities are integrated into courses, the more meaningful their contribution to low-achieving students will be, enhancing their ability to cope with academic studies.

Table 1
Grouping of Students by Pre-SQG Scores

|  |  |  |  |
| --- | --- | --- | --- |
| Percent | Frequency | Range of scores | Group |
| 13.5 | 23 | 0–54 | Low scores |
| 23.4 | 40 | 55–69 | Intermed. scores |
| 36.3 | 62 | 70–84 | Good scores |
| 26.9 | 46 | 85–100 | Very good scores |
| 100.0 | 171 | 4 | Total |

Table 2
ANOVA Test for Inter-Group Variance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables examined | Group | Mean | SD | F | Sig. |
| Overall score before SQG | Low scores | 41.17 | 11.05 | 466.37 | .000 |
| Intermed. scores | 62.63 | 3.95 |
| Good scores | 76.35 | 4.32 |
| High scores | 89.52 | 3.04 |
| Total | 71.95 | 16.36 |
| Overall score after SQG  | Low scores | 49.87 | 15.34 | 87.41 | .000 |
| Intermed. scores | 68.20 | 8.40 |
| Good scores | 77.74 | 7.48 |
| High scores | 85.57 | 7.41 |
| Total | 73.87 | 14.47 |
| Higher-order-thinking question scores before SQG | Low scores | 15.22 | 18.06 | 49.75 | .000 |
| Intermed. scores | 29.75 | 22.31 |
| Good scores | 52.58 | 21.95 |
| High scores | 73.15 | 21.38 |
| Total | 47.75 | 29.20 |
| Higher-order-thinking question scores after SQG | Low scores | 31.52 | 26.35 | 34.95 | .000 |
| Intermed. scores | 51.50 | 21.19 |
| Good scores | 66.77 | 17.27 |
| High scores | 79.89 | 18.72 |
| Total | 61.99 | 25.28 |

Table 3
Overall Score of Student Groups, Before and After SQG

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *p* | *df* | *t* | Overall score after SQG (SD) | Overall score before SQG (SD) | Group |
| .000\*\* | 22 | 4.58 | 49.87(8.40) | 41.17 (11.05) | Low scores |
| .000\*\* | 39 | 3.98 | 68.20(17.99) | 62.63 (4.32) | Interm. scores |
| .144 | 61 | 1.48 | 77.74(7.48) | 76.35(18.72) | Good scores  |
| .201 | 45 |  1.06 | 88.89(6.41) | 89.52 (3.05) | Very good scores |

Table 4
Score on Higher-Order Questions, Before and After SQG

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *p* | *df* | *t* | Score after SQG (SD) | Score before SQG (SD) | Group |
| .000\*\* | 22 | 3.65 | 31.52(19.41) | 15.22 (11.06) | Low scores |
| .000\*\* | 39 | 5.51 | 51.50(21.18) | 29.75 (22.33) | Intermed. scores |
| .000\*\* | 61 | 5.34 | 66.77(17.27) | 52.58(21.94) | Good scores  |
| .098 | 45 |  1.69 | 79.89(18.72) | 73.15 (21.37) | Very good scores |

**Appendix: The SQG Model**

1. In the fifth lesson of the second semester, the students were presented with examples of questions at various orders of thinking on a topic that had already been covered. Although their education courses had already familiarized them with the concepts pertaining to the types of questions and Bloom’s taxonomy, they had little experience in classifying questions. To simplify matters, they were presented with a two-group taxonomy: basic knowledge and memorization questions, and higher-order-thinking questions that included all other types of questions such as comprehension, application, and synthesis. The activity lasted around thirty minutes and at its end the students were given a homework assignment to be completed in pairs. The exercise, a course requirement, included generating three questions on the topic taught in class and at least two transformation questions. The students were required to upload the questions to the course website within a week and to answer and comment on another pair’s questions.
2. In the seventh lesson, examples of student questions were presented in class and were discussed, addressing the level and clarity of the questions as well as possible answers. The class activity lasted around forty minutes.
3. The final lesson in the second semester was devoted in its entirety to generating and answering questions by the students. At the beginning of the lesson, it was emphasized to students that the activities in the lesson would help them to summarize and organize the material and would result in the construction of a question bank that would help them to review for the exam. Moreover, the sequence and nature of the activities, as detailed below, was briefly explained in advance:
* The teacher divided the class into four or five groups comprised of three or four students each, depending on the size of the class. Each group was heterogeneous in regard to its achievements in the first semester, including students who earned high scores and mediocre or poor scores at that time.
* Each group was given one main topic among the topics studied during the second semester and was asked to generate five questions about it, including at least three transformation or higher-order-thinking questions. They were given forty minutes to generate the questions and were allowed to use materials from lectures, the course website, digital books, and various websites as aids.
* The teacher facilitated the groups during the SQG activity, mostly in the sense of helping them to generate the transformation questions and encouraging less-active students to participate. The group uploaded the questions to the course website only after the teacher approved them.
* When the forty minutes allotted to the SQG activity finished, each group received another group’s questions and spent around thirty minutes answering them and also commenting on their level and clarity. The answers and the comments were then given to the group that generated the questions; the group then checked the answers to, and read the comments on, their questions.
* When the activity was over, a bank was created of around twenty-five questions covering all of the course topics in the second semester. This was uploaded to the course website. Some 60 percent of the questions were of the higher-order-thinking type.

In sum, the sequence of question-generation, question-answering, and peer-assessment activities was as follows: (1) class discussion of the types of questions and their classification; (2) a homework assignment to generate, answer, and assess questions; (3) class discussion of the homework assignment; (4) a group activity in class to generate, answer, and peer-assess the questions; (5) creation of a question bank.