Student-Generated Questions:
Developing Mathematical Competence through Online Assessment

Sima Caspari-Sadeghi, Brigitte Forster-Heinlein, Jutta Maegdefrau, and Lena Bachl

University of Passau

This action research study presents the findings of using a formative assessment strategy in an online mathematic course during the world-wide outbreak of Covid-19 at the University of Passau, Germany. The main goals of this study were (a) to enhance students' self-regulated learning by shifting the direction of assessment from instructors to the students, (b) to promote deep active learning in mathematics. Students were required to conduct self-regulated learning on a selected topic. They were encouraged to formulate two multiple-choice questions (MCQs) and pose them after each presentation in an online course. The effectiveness of Student-generated Questions (SGQs) as a learning strategy was measured in terms of (a) students' engagement, and (b) learning outcomes. While evidence on students' engagement was gathered through an online questionnaire survey, the learning outcomes were measured by analyzing the quality of SGQs. Results indicated that authoring questions, though leading to a higher students' engagement with the materials, could be quite challenging for students and did not lead to higher achievement. The authors provide some suggestions for improving the process through regular uses of digital technologies such as PeerWise.

INTRODUCTION

Questioning is rarely used as a knowledge-seeking method. Those who ask questions—teachers, texts, tests—are not seeking knowledge; those who would seek knowledge—students—do not ask questions (Dillon 1988, p. 197).

Questions are at the heart of any scientific learning; a new inquiry begins with a question, and the consequent search for answers advances our understanding of the phenomenon. Cognitive science argues convincingly that questioning acts as a psychological tool for reflecting, critical thinking, scaffolding ideas, and social interaction. Using questions as a teaching strategy dates back to Socrates. His dialectical method is still one of the most powerful tools for promoting critical thinking. Instead of directly answering students' questions, Socrates entered into an argumentative dialogue based on focused questioning, which led the students to find the answers themselves (Paul and Elder, 2008). The so-called "Pedagogy of Inquiry" (Pagowsky, 2015) is also a teaching method that involves student-centered classroom questioning and investigation in order to encourage metacognitive thought processes, discussion, and collaboration.

Mathematics is inherently an inquisitive discipline which evolves around questions and problems. In a typical classroom, teachers retain control of asking questions: the questions are initiated by the teachers and students take their turn to answer. There are opportunities where the teacher invites students to ask questions. However, when the teacher is the one who constructs the most interesting questions and problems, students become dependent upon the teacher to catalyze inquiry (Bowker, 2010). To facilitate development of mathematical competence, teachers should create effective learning environments and encourage students to ask relevant and scientifically sound questions (Foster, 2011; Penick, Crow, & Bonnsteter, 1996).

This action research aimed at moving beyond teacher questioning by making students actively involved in the process of knowledge construction via generating questions. The main objective was to encourage self-regulated learning. Self-regulated learning is defined as the ability to pursue and persist in learning, and to organize one's own learning. It includes effective management of time and information, both individually and in groups (Redecker, 2013). Students were required to conduct research on assigned topics, summarize their findings, formulate two MCQs, and present them in an online mathematic course. Student-Generated Questions (SGQs) served the dual goals of (a) maximizing intellectual involvement with content and making sure the audience give a focused attention to the presentation instead of passive listening, (b) empowering the students to develop the competence and confidence in formulating quality questions in mathematics.

The following broad questions were addressed:

R.Q.1: Does SGQ strategy enhance students' motivation and participation in learning Mathematics?

R.Q.2: Is there a relationship between the quality of SGQ and their mathematical competence?

The Covid-19 pandemic was an impetus for shifting to online assessment in this project. We wanted to embed formative, online assessment to get continuous feedback on students' learning.

LITERATURE REVIEW

Questioning may serve different functions, depending on whether it is being used as a "teaching strategy" or a "learning strategy". Teachers ask questions mainly to (a) monitor progress in students' understanding; (b) identify gaps in knowledge; (c) diagnose misunderstandings/misconceptions; (d) stimulate the recall of prior knowledge; and (e) generate peer-to-peer interaction (Tofade et al., 2013; Bell & Cowie, 2001). On the other hand, students' questions have the potential to (a) direct their learning and drive knowledge construction; (b) increase their motivation and interest in a topic by arousing their epistemic curiosity; (c) foster the quality of discourse and classroom talk; and (d) help them to self-evaluate and monitor their understanding (Chin and Osborn, 2008).

Several studies noted the significance of questioning skills in scientific literacy. Earlier studies on SGQs aimed primarily at raising awareness about the dearth of SGQs in the classroom (Corey, 1940); its negative consequences, such as reducing students to

"passive respondents" instead of active initiators of questions (Tizard et al., 1983); or reinforcing gender inequality, i.e. female students pose fewer questions compared to their peer males (Pearson and West, 1991).

Background knowledge seems to play an important role in the quality of SGQs. McQueen et al. (2014) found that students with lower levels of prior knowledge failed to engage critically with the content and produced lower-quality questions.

Some studies focused on judging the quality of the questions. In addition to using Bloom's Taxonomy for measuring the cognitive demand and difficulty level of the SGQs (Bates et al., 2014), some tried to develop a "questioning rubric" addressing the quality of the questions (Guthrie et al., 2007), while others used criteria such as correctness and clarity (Papinczak et al., 2012) or the plausibility of distractors (Purchase et al., 2010).

A number of empirical studies examined the influence of practising SGQs on some measurable student performance. In a quasi-experimental study, Shakurnia et al. (2018) examined the impact of writing SGQs on immunology test scores. The results indicated a positive impact, with the treatment group who practised question writing gaining a higher score on an achievement test.

King (1994) studied the effect of SGQs on students' problem-solving skills. The results indicated a positive impact only for the group tha received guidance and mentoring on question formulation. In a similar study, Byun et al. (2014) investigated the impact of questioning strategies on solving ill-structured problems. They found teacher-initiated questions were more beneficial for developing problem-solving skills. It was argued that the mere construction of questions does not automatically lead to a higher outcome. Unless students get explicit support and put substantial effort into creating high-level questions, the process may not contribute to learning gain.

RESEARCH CONTEXT AND METHODOLOGY
Action Research
This action research was conducted within the framework of SKILL.de, a project aiming at digitally enhancing higher education, at the University of Passau, Germany (2019-2023). It was a joint collaboration between an expert in action research (called researcher) and two teaching staff, a Professor and her Teaching Assistant, in the department of mathematics (called instructors). The main author is responsible for faculty professional development in evidence-based evaluation. The goal is to ensure that any adopted intervention, e.g., the use of digital technology or the imlementation of innovative teaching and learning strategies, is backed by evidence that is collected and analyzed by the practitioners themselves. The researcher offered coaching, mentoring, and training in action research to enable the teaching staff to identify a problem, plan an intervention, implement it, collect and evaluate evidence, and systematically reflect on their learning. It could be considered as practical action research, in which teachers are the researchers of their own practices (Mills, 2018; Stenhaus, 1975). This project began with an observation that the instructor experienced frequently in her courses: during the course, students are asked if they have any questions. There are rarely any. However, at the last session, students suddenly come up with several questions. Diagnosis of this problem shaped the subsequent intervention.

Procedure
This study was conducted in an applied mathematics course during the Covid-19 pandemic (March-September, 2020). The course was run in an online synchronous mode (via ZOOM) with the further support of Stud.IP (the University's Learning Management System). The course was taught jointly by a professor and her teaching assistant who had recently graduated. The participants were ten bachelor and two master students with no prior experience in online instruction.

The main goal was to develop "learning to learn" as one of the "*key competences for life-long learning*" (European Commission, 2012). To achieve this goal, the students were required to select a related topic in mathematics, conduct inquiry on it, and present their summary and findings to the classroom. Self-directed learning was supported by the instructors, who recommended literature, answered questions, etc.

Although transferring the responsibility of learning and teaching (peer-instruction) is practiced in many disciplines, the responsibility for assessment remains mostly with the teacher. Therefore, it was decided that self-regulated learning skill should be also embedded in the assessment process, which would be conducted by the students themselves. Each student was required to formulate two MCQs and present them at the end of their talk. They were instructed verbally about the quality criteria, such as clarity, content coverage, relevance, and plausible distractors. Every session, two students presented their topics and evaluated the understanding of their audience with their SGQs. The class had two minutes of thinking-time for each question. They voted via Stud.IP, and the results were immediately available in a graphical chart with percentages. There was a short time for explanation and discussion after each question.

ANALYSIS
Questionnaire on students' attitude
Students' perceptions of the value of the SGQs strategy were assessed through a brief self-reported online questionnaire administered at the end of the course. It contained two questions (Q1. Impact of SGQs on their involvement, Q2. If SGQs caused excessive pressure on them). Students could answer each question on a five-point Likert scale (1=strongly disagree, 5=strongly agree). Findings revealed that students had a positive attitude towards the experience, with 88% of the participants reporting that it contributed greatly to their focused attention and engagement with the course content. More than 62% believed writing questions didn't impose excessive work or pressure on them.

SGQs Analysis
The collected SGQs were analyzed based on a two-dimensional rubric which addressed (a) the overall quality of a question, and (b) the cognitive demand involved in a question. Each dimension has several levels of performance, with traits specific to each level.

Overall Questions quality
The overall quality of the questions, including stems and distractors, was measured on a rating scale of (1= poor, 2=good, 3= excellent) based on three criteria: content coverage, relevance/clarity, and plausibility. For example, if a question was formulated

clearly and unambiguously, had plausible distractors, and related significantly to the topic, it was rated as 3– excellent (see appendix for a sample of SGQs).

Cognitive demand

Furthermore, each question was judged in terms of its difficulty as measured by the cognitive level involved in answering the question. The following table, Bloom's Taxonomy, shows the six levels in cognitive learning (Anderson & Krathwohl, 2001; Bloom, 1956).

Table 1. Bloom's Taxonomy of Cognitive Levels		
Cognitive domain	Cognitive level	Action required
Knowledge (remembering)	Low	Recognition, recall, name, list
Comprehension	Low	Describe, explain, summarize, visualize
Application	Low	Use, practice, solve, manipulate
Analysis	High	Compare, deduce, analyze, infer
Synthesis	High	Synthesize, devise, design, construct, plan
Evaluation	High	Judge, criticize, estimate, justify, defend

Inter-rater Reliability

The researcher provided the practitioners (instructors) with training on the rubric for judging the quality of the SGQs. The professor and her assistant rated the quality of each question based on cognitive demand and overall quality. Kappa Measure of Agreement was calculated to check the consistency of the classification. The interrater reliability of Kappa= 0.76 (with the significance of $p < .000$) could be partly due to the different expertise levels of the raters.

RESULTS

Figure 1 illustrates the percentage and quality of SGQs at each of Bloom's cognitive levels. The majority of the questions generated by students (66%) were classified at the lowest category (remembering), 25% at level 2 (understanding), and less than 10% at level 3 (application) of Bloom's taxonomy. The lack of plausible distractors, ambiguous wording, and incorrect assumptions in the question stem were the major deficiencies in the SGQs. None of the SGQs were at the higher-order levels, such as analysis, synthesis, or evaluation/creativity. No significant positive correlation could be established between the quality of the SGQs and the students' academic level of achievement on the final exam.
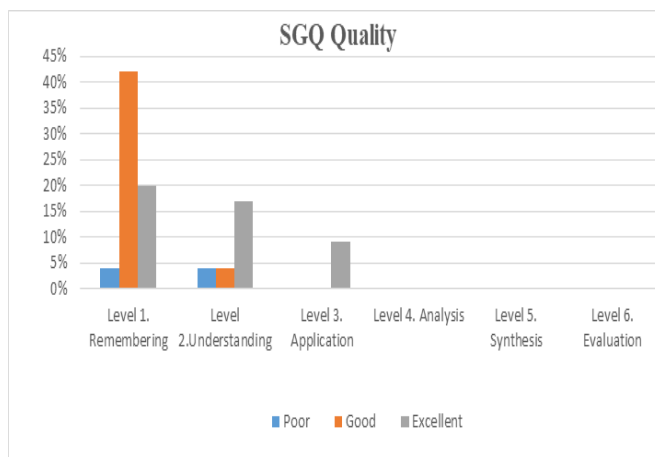
DISCUSSION

There might be a number of tentative explanations for these findings. First, Bottomley and Denny (2011) suggested that such results are to be expected, since this was likely the first time these students were asked to write their own questions systematically. Development of appropriately aligned MCQs is not an easy or trivial task. In a large-scale review of university biology courses across the United States, it turned out that about 90% of MCQs generated by instructors targeted the lowest two levels of the Bloom's taxonomy (Momsen et al., 2010). Zheng et al. (2008) used the same mapping procedure to analyze questions produced at the university level and reported similar results: a high proportion of questions fell in the lowest two levels, "remembering" and "understanding". Second, MCQs (or any selected-response items) are often criticized for their inadequacy in measuring high-level knowledge and understanding (Hickson & Reed, 2009). This view was asserted by the instructor: "A lot of mathematical concepts and processes, i.e. derivations and proofs, cannot be adequately assessed by MCQs" (B. Forster-Heinlein, personal communication, 2020). Another explanation could be the "low-stakes" nature of the task; since the students did not get any specific mark as a reward, it was difficult to ensure if they put serious, thoughtful effort into it.

Though there is no substantial evidence of deep learning, based on the students' survey results, it can be concluded that authoring questions presented a richer learning experience and engagement in the course material for them.

CONCLUSION: CLOSING THE ACTION RESEARCH CYCLE

> *Purposeful inquiry does not happen spontaneously—it must be learned.* (Baird, 1990, p. 184)

Generating well-crafted questions is a creative act, and is at the heart of what doing science is all about (Chin and Osborn, 2008). Students-initiated questions open a window to the mind of the students: they indicate what counts as significant for the students, what they understood, misunderstood or missed altogether. In this empirical study, we tried to foster a "culture of inquisitiveness" and self-regulated learning by encouraging development of SGQs in mathematics.

The findings of this study were consistent with some previous literature (e.g., Bottomley and Denny, 2011), that although students appreciated the opportunity to create questions and considered it as a valuable activity, it did not lead to higher-order thinking or improve their mathematical competence significantly. However, there are limitations in such small-scale case studies as this, with its small sample size and limited data, which mandates the findings be treated with caution. Therefore, the authors make no claim on generalizability beyond the context of this project.

The cycle of Action Research is completed when its practitioners reflect on what they learned from the project and plan for the future improvement of such practices (Bruce and Flynn, 2019; Tsang, Annetta, 2009). All authors were actively involved as reflective practitioners who continually tried to observe the impact of introducing a "Learning Assessment Technique" addressing students' learning outcomes in an online environment. Regular meetings and discussions during the whole process of data collection, analysis, and interpretation by authors from different backgrounds (Mathematics, Education and Learning sciences) led

to a deeper, more integrated and comprehensive grasp of the challenges that students were facing. Reflection on these findings led to the conclusion that this project could be improved by harnessing the affordances of digital technology as well as the instructors' support.

White (1977) suggested that the ability to formulate questions is a skill which needs to be taught rather than left to chance. The support of an instructor, which includes coaching and feedback, plays a key role in scaffolding the writing process. Successful strategies might include "explicitly teaching" about formulating quality MCQs (stem, options, key, distractors), providing some examples of "best practices", discussing the weaknesses and strengths of sample questions, and modeling higher-order questions (those questions that require critical thinking, need evaluative judgement, or look for cause and effect). Jobs et al., (2013) showed that by providing students with MCQ-writing manuals, instructors managed to improve SGQs proactively. However, for most faculty members, such an endeavor is both time-consuming and labor-intensive.

Employing digital technology is another strategy for enhancing both the efficiency (time- and workload-saving) and effectiveness (learning improvement) of such efforts. PeerWise (Kelley et al., 2019) is a freely available online tool, developed in the Department of Computer Science at the University of Auckland, which allows students to author, answer, and discuss content-related questions (https://peerwise.cs.auckland.ac.nz/). In addition to writing their own MCQs, students can answer, review, and rate their peer's questions, provide comment and feedback, and even follow other contributors. Several studies have assessed the impact of PeerWise on student engagement and learning (Kay et al., 2020). Most report a positive effect, both in terms of student engagement and mastery of learning outcomes as measured by the correlation between system usage and achievement test scores (Shield et al., 2020).

PeerWise should not be viewed merely as a "repository of SGQs", but as a vehicle for shifting the locus of control back to students, encouraging them to take more responsibility for meaningful learning and sustainable assessment, which benefits them beyond their academic lives. Further research is needed in order to fully investigate the contribution and suitability of SGQs in the development of students' mathematical competence.

ACKNOWLEDGEMENT

REFERENCES
Anderson, L.W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete edition). New York: Longman.

Baird, J.R. (1990). *Metacognition, purposeful inquiry and conceptual change*. In E. Hegarty-Hazel (Ed.), The student laboratory and the science curriculum (pp. 183–200). London:Rutledge.

Bates, S. P., Galloway, R. K., Riise, J., & Homer, D. (2014). *Assessing* the *quality* of a *student-generated question repository*. *Physical Review* Special Topics. *Physics Education Research*, *10*(020105), 1-11.

Bloom, B.S., Engelhart, M.B., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York: Longmans Green.

Bottomley S, & Denny P. (2011). A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education, 39*(5), 352-61.

Bowker, M. (2010). Teaching students to ask questions instead of answering them. *Thought & Action, 20,* 127-134.

Bruce, C., & Flynn, T. (2019). Action research as professional learning for educators. In *The Wiley Handbook of Action Research in Education*, 273-294.

Byun, H., Lee, J., & Cerreto, F.A. (2014). Relative effects of three questioning strategies in ill-structured, small group problem solving. *Instructional Science, 42*(2), 229-250.

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education, 44*(1), 1-39.

Corey, S. M. (1940). The teachers out-talk the pupils. *The School Review, 48*(10), 745–52.

Dillon, J.T. (1988). The remedial status of student questioning. *Journal of Curriculum Studies, 20,* 197–210.

European Commission. (2012). Assessment of Key Competences in initial education and training: Policy Guidance. *Staff Working Document. Accompanying the Communication from the Commission on Rethinking Education: Investing in skills for better socio-economic outcomes*. SWD 371.

Foster, C. (2011). Student-generated questions in mathematics teaching. *Mathematics Teacher, 105*(1), 26–31.

Guthrie, J. T., McRae, A., & Klauda, S. L. (2007). Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist, 42,* 237-250.

Hickson, S., & Reed, W.R. (2009). Do constructed-response and multiple-choice questions measure the same thing? Department of Economics. Working Paper Series.

Kay, A., Hardy, J., & Galloway, K. (2020). Student use of PeerWise: A multi-institutional, multidisciplinary evaluation. *British Journal of Educational Technology, 51*(1), 23-35.

Kelley, MR, Chapman-Orr, EK, Calkins, S & Lemke, RJ (2019), 'Generation and Retrieval Practice Effects in the Classroom Using PeerWise', *Teaching of Psychology, 46*(2), 121-126.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31,* 338–368.

McQueen, H. A., Shields, C., Finnegan, D.J., Higham, J., & Simmen, M. W. (2014). PeerWise provides significant academic benefits to biological science students *across diverse learning tasks, but with minimal instructor intervention, Biochemistry and Molecular Biology Education, 42*(5), 371–381.

Mills, G.E. (2018). *Action research: A guide for the teacher researcher.* 6th edition, Pearson.

Momsen, J.L., Long, T.M., Wyse, S.A., & Ebert-May, D., (2010) Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE-Life Sciences Education, 9,* 435-440.

Pagowsky, N. (2015). A Pedagogy of Inquiry. *Communications in Information Literacy, 9*(2), 136-144.

Papinczak, T, Peterson, R, Babri, AS, Ward, K, Kippers, V & Wilkinson, D. (2012). Using student-generated questions for stu-

dent-centred assessment. *Assessment & Evaluation in Higher Education, 37*(4), 439–52.

Paul, R. & Elder, L. (2008). Critical thinking: the art of Socratic questioning, part III. *Journal of Developmental Education,31*(3), 34-35.

Pearson, J. C., & West, R. (1991). An initial investigation of the effects of gender on student questions in the classroom: Developing a descriptive base. *Communication Education, 40,* 22-32.

Penick, J. E., Crow, L.W., & Bonnstetter, R. J. (1996). Questions are the answer: A logical questioning strategy for any topic. *The Science Teacher, 63,* 27-29.

Purchase, H., Hamer, J., Denny, P. & Luxton-Reilly, A. (2010). The quality of a PeerWise MCQ repository, in *'Proceedings of the Twelfth Australasian Conference on Computing Education-Volume 103',* Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 137-146

Redecker, C. (2013). The use of ICT for the assessment of key competences: Institute for prospective and technological studies, Joint Research Centre.

Shakurnia, A. Aslami, M. & Bijanzadeh, M. (2018). The effect of question generation activity on students' learning and per-

ception. *Journal of Advances in Medical Education and Professionalism, 6*(2),70-77.

Shields, T., Gray, G., & Ryan, B. J. (2020). Asynchronous assistance: A social network analysis of influencing peer interactions in PeerWise. *European Journal of Mathematics and Science Education, 1*(1), 43-52.

Tizard, B., Hughes, M., Carmichael, H., & Pinkerton, G. (1983). *Children's questions and adults' answers. Journal* of *Child Psychology* and *Psychiatry, 24,* 269–81.

Tofade, T., Elsner, J., & Haines, S.T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education, 77*(7).

Tsang, Annetta (2009). Personal Reflection: Reflective Learning as a Student and an Educator: Connecting the Scholarship of Teaching and Learning. *International Journal for the Scholarship of Teaching and Learning, 3*(2), 29.

White, R.T. (1977). An overlooked objective. *Australian Science Teachers' Journal, 23*(2), 124-125.

Zheng, A.Y., Lawhorn, J.K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy: Debunks the "MCAT myth". *Science, 319,* 414-415.

APPENDIX. SAMPLE OF SGQS

1. Convert the digital number 67.25 into the octal numeral system. Which answer gives the correct representation?

    a- a.34
    b- 103.2
    c- 1000011.01
    d- 1003.1

2. Which of the following was not part of the chain reaction that had been triggered by the unsafe cast?

    a- The SRIs one after the other turned off.
    b- The OCB interpreted the meaningless bit pattern as correct measurements.
    c- The main rocket engine was temporarily turned off.
    d- The system threw a hardware exception.

3. In which of the following cases the Nash equilibrium is uniquely defined?

    a- F is quasi-convex.
    b- F is pseudo-monotone.
    c- F is uniformly monotone.

4. Are there one or several Nash equilibria in the game "rock-paper-scissors"?

    a- Yes, in the points (scissors, scissors), (rock, rock), (paper, paper).
    b- No, there is no Nash equilibrium.
    c- Yes, in the points (scissors, rock), (rock, paper), (paper, scissors).