Assessment through mathematical problem-posing

Rogier Bos and Rebecca Kuijpers

Utrecht University, the Netherlands; r.d.bos@uu.nl

This study concerns problem-posing as a means of assessment in upper-secondary mathematics education. The open character of problem-posing as a task, allows students to show their creativity but makes it difficult to control the focus on the learning goals. Problem-posing can be structured by adding an initial problem to the prompt. We aim to investigate how this form of structuring affects the resulting problems and the extent to which they reveal students' thinking and knowledge with respect to learning goals. In line with previous research on assessment through problem-posing by Kwek and by Mishra and Iyer, we classify the complexity of the problems. Additionally, we analyze whether the problems address the learning goals and are solvable. The main outcome is that structuring the problem-posing prompt is more suitable for assessment since the resulting problems align better with the learning goals and reveal more of the qualities and misunderstandings of the students.

Keywords: Assessment, Mathematics education, Problem-posing.

Introduction

Problem-posing is a teaching approach where students are invited to create or reformulate a problem rather than solve a given one. It has been implemented in some national curriculums and is part of the U.S. Principles and Standards for School Mathematics of the NCTM. Research on problem-posing has been ongoing since at least the ’90s (Silver, 1994; Stoyanova, 1997); see, e.g., the two recent reviews by Baumanns & Rot (2020; 2021). Reasons to teach problem-posing are, among others, that it invites students to analyze situations and that it fosters their creativity (Baumanns & Rot, 2020)

In analogy to the well-known distinction between teaching problem-solving and teaching *through* problem-solving, one can distinguish between assessing problem-posing and assessing *through* problem-posing. The former has been studied extensively (e.g., Silver & Cai, 2005). However, research on assessing through problem-posing is limited to our knowledge (Kwek, 2015; Mishra & Iyer, 2015). If students are taught through problem-posing, as advocated Zhang and Cai (2021), assessing through problem-posing would improve constructive alignment (Biggs & Tang, 2011). The question is: how can we assess a student's learning, knowledge, and skills by considering the problems they pose? The open character of problem-posing tasks seems to prevent teachers from focusing the outcome on the learning goals. Hence the handed-in problems may not allow teachers to assess what they intend to assess. Moreover, the problems students pose may not display the level at which the teacher intended to assess. However, the same open character also allows teachers to assess students’ creativity (Baumanns & Rot, 2021). It may allow students to show what they can do beyond what the teacher might envision, and that could be an attractive property for assessment.

In this study, we investigated how structuring the problem-posing prompt, i.e., providing an initial problem as part of the prompt, might help teachers to nudge the problem-posing in the direction of the desired learning goals and the desired level. It is a trade-off: by structuring the problem the task will be less open, but more focused. Does that hamper creativity? Does it lead to a more informative assessment? By comparing two groups of 10th graders that pose problems, with or without an initial problem, we aim to study how problem-posing can be used for assessment, and whether adding an initial problem improves assessment.

Theoretical background

Stoyanova and Ellerton (1996) differentiate problem-posing situations as free, semi-structured, or structured. The problem-posing starts from a provided real-life or artificial situation. If there are no further restrictions, the situation is called *free*. In a semi-structured situation, the problem should require prescribed mathematical concepts or skills. In a structured situation, students are provided with an initial problem, after which they are invited to pose more problems about the same situation. In the latter case, mathematical concepts and skills are suggested by the initial problem. Baumanns and Rott (2020) often take the two types free and semi-structured together, and so shall we in this article, using the label *unstructured situation*.

We found two papers that explicitly address problem-posing as a means of assessing using semi-structured prompts: Kwek (2015) and Mishra & Iyer (2015). Kwek (2015) introduced rubrics to classify the complexity of the problems that students pose. Table 1 shows a shortened version of these rubrics. Kwek analyzes a set of problems posed by 7th and 9th graders. For grade 9, 78% of the problems are solvable, 67% are of low complexity, 30% are of moderate complexity, and 3% are of high complexity. Students were invited to discuss each other's problems and decide whether these were interesting and challenging. The grade 9 students found 58% of the problems interesting and 50% of the problems challenging. The grade 9 students showed appreciation for problems with strong mathematical content. Kwek concludes that both cognitive factors, like thinking processes and understanding, and affective factors were revealed through classroom problem-posing, making it a suitable assessment activity. However, we believe it could be of additional interest to focus on how the problems reveal students' misunderstanding and to see whether problems cover learning goals.

Mishra and Iyer analyze problems posed as part of a computer science course. Similarly to Kwek, they classify the complexity of the problems based on rubrics: 39% low, 51% medium, and 10% high. 85% of the advanced students who scored high on a classical assessment still produced a problem of medium to low complexity. This indicates that students are not necessarily challenged to perform to their highest ability by a problem-posing assignment. Mishra and Iyer also track which learning goals concerning computational thinking are covered by the problems. They find that some learning goals are better addressed than others, ranging from some goals only covered by 8% of the problems to others by 96%. Missing out on certain learning goals in an assessment can be problematic. In this paper, we study whether providing structured prompts improves such coverage of the learning goals.

In line with these papers, we propose, when assessing through problem-solving, to take into account the solvability of the problem, the complexity of the problem, and the extent to which it covers the learning goals. Solvability is a measure of correctness: if students produce a problem that cannot be solved, this influences the assessment in the same way an incorrect answer influences a traditional assessment. Moreover, it makes sense to take the complexity into account. A problem-posing task has a degree of freedom that could best be compared to the difference between a simple correct answer and an impressive correct answer: complexity is a quantity that allows one to capture this dimension.

As part of the study that we report on in this paper, we compare structured with unstructured problem-posing as a form of assessment. Our research question is: how do these types of situations and prompts for problem-posing contribute to assessment? Is one type more suitable than the other? We restrict ourselves to high-achieving 10th-graders and the subject of probability, but later discuss whether the results might extend beyond these specifics.

Table 1. Complexity of a posed problem; adapted from Kwek (2015)

	Low complexity	Moderate complexity	High complexity
Description	The problem typically specifies what the solver is to do, which is often to carry out some procedure that can be performed mechanically.	Solving the problem involves more flexible thinking and choice among alternatives. It requires going beyond routine approaches or using multiple steps.	High-complexity problems make heavy demands on solvers, who must engage in more abstract reasoning, planning, analysis, judgment, and creative thought.
Cognitive demand	• Recall or recognize a fact, term, or property • Perform a specified routine of steps • Retrieve information from a graph, table, or figure	• Represent a situation mathematically in more than one way • Justify steps in a solution process • Interpret a visual representation • Extend a pattern • Interpret a simple argument	• Perform a non-routine procedure having multiple steps and multiple decision points • Generalize a pattern • Explain and justify a solution to a problem • Provide a mathematical justification

Method

The study was performed in the context of Mathematics D Online, a Dutch nationwide hybrid (mixed online and onsite) course on advanced mathematics for high-achieving secondary school students. As part of this program, students were invited to hand in answers to a weekly set of tasks. 275 students aged 15 to 16 were enrolled in the course in 2022/2023. We had a sample of 20 students, which we found sufficient, based on previous studies on problem posing with similar data collection (Kwek, 2015; Stoyanova, 1997). The sample was not random, but based on student's positive replies to a request to participate: a convenience sample. We replaced four of the hand-in tasks on probability with problem-posing tasks. Each task consisted of a context and a prompt. For both structured and unstructured tasks, the context was identical, but the prompt differed (see Table 2).

The handed-in problems and accompanying answer models were analyzed and coded for coverage of the learning goals, complexity, and solvability by the second author. The first author performed a second coding. The coverage of learning goals was determined by comparing it to a list of learning goals, representing the material the students were working on. The complexity was coded using an extended version of Table 1. The solvability was determined by carefully examining the problem and the answer model, taking into account that the problem needs to be clear about what needs to be solved and provide enough information to do so, and that it needs to be mathematically correct and consistent. Next, these results were statistically analyzed with suitable tests to allow comparison of structured and unstructured prompts.

Table 2. Examples of problem posing tasks

<table>
<tr><th rowspan="2">Context</th><th colspan="2">Prompt</th></tr>
<tr><th>Structured</th><th>Unstructured</th></tr>
<tr><td>A random variable X has the following distribution:
<table><tr><td>X</td><td>0</td><td>1</td><td>2</td></tr><tr><td>$P(X = x)$</td><td>$\frac{1}{2} - \frac{p}{2}$</td><td>$p$</td><td>$\frac{1}{2} - \frac{p}{2}$</td></tr></table>with $0 \leq p \leq 1$.</td><td>a. Compute the standard deviation in terms of p.
b. Pose two more problems on this distribution. Also make the answer model.</td><td>Pose three problems on this distribution. Also make the answer model.</td></tr>
<tr><td>A factory produces blue, green, and red soaps. The weight of a soap is normally distributed, with $\mu = 100g$ and $\sigma = 3g$ for blue soaps, $\mu = 120g$ and $\sigma = 4g$ for red soaps and $\mu = 80g$ and $\sigma = 3g$ for green soaps. The volume of a soap is normally distributed, where $\mu = 0{,}2l$ and $\sigma = 0{,}002l$ for blue soaps, $\mu = 0{,}25l$ and $\sigma = 0{,}003l$ for red soaps and $\mu = 0{,}18l$ and $\sigma = 0{,}003l$ for green soaps. The factory sells blue soaps for €1,-, red soaps for €1,50 ,and green soaps for €0,85. The number of soaps sold per day is normally distributed, with $\mu = 40$ and $\sigma = 3$ for blue soaps, $\mu = 35$ and $\sigma = 2{,}5$ for red soaps and $\mu = 40$ and $\sigma = 2{,}5$ for green soaps.</td><td>a. Compute the probability that the volume of a blue soap is less than 0,24l or more than 0,26l.
b. Pose two more problems on this distribution. Also, make the answer model.</td><td>Pose three problems on this distribution. Also make the answer model.</td></tr>
</table>

Results

Nine students handed in 33 problems from structured prompts, and 11 students handed in 53 from unstructured prompts. All problems were coded by the second author. For assessing interrater agreement, 1/3 of the problems were coded by the first author. The codes on solvability and learning goals were identical. Before discussion, there was a Cohen's kappa of 0.66 on the coding of complexity. This was mainly due to confusion about whether a multistep problem should be coded as low or moderate complexity. The raters decided to add the distinction "routine versus non-routine" to the complexity matrix. This decision gave clarity on most differences, leading to a Cohen's kappa of 0.96.

Sample problems

We discuss three problems from our sample here. The first problem was constructed as a response to the first context. The posed problem was: Compute the standard deviation when $p = 1/2$. The problem was coded as having a low complexity because computing the standard deviation is a routine procedure for these students. The problem is solvable and covers a learning goal, namely computing the standard deviation from a probability distribution. Hence the problem posed and the answer model allow us to assess the student's progress for this learning goal, but only on a reproductive level.

The second problem was constructed as a response to the second context. The posed problem was: Compute the probability that the gain of the factory is more than €36, just from the green soap. It was coded as having a moderate complexity because it is not a routine problem. While the computation is a standard one, the solver first has to realize that the answer lies in the number of soaps sold, not the price of the soap. The covered learning goal is: to compute probabilities using the normal distribution, where the average value, standard deviation, and the boundaries are given. The problem was formulated with clarity and is solvable. Hence, it allows us to assess the students on this goal and also conclude that a level of flexibility and creativity was achieved.

The third problem was also formulated in the second context and consists of two parts. Part 1: Which color of soap has a higher probability of being sold less than 39 times per day? Part 2: what are the differences in those probabilities per color? This problem, too, was coded as having a moderate complexity, because the problem, apart from computing probabilities, involves comparing these probabilities. The problem covers the same learning goal as the second problem. The problem is not properly solvable, because of an accidentally too open formulation: the student did not specify how the difference between probabilities should be expressed. While it is unusual to rate the complexity of unsolvable problems (see for example Silver & Cai, 1996), we chose to do so, because with all these problems the intention of the author of the problem was clear, also from their answer model. The formulation of the problem reveals a gap in knowledge about how probabilities should be compared.

Problems covering the learning goals

For each problem posed we analyzed whether it covered at least one learning goal. Of the structured tasks, 97.0% addressed at least one learning goal, whereas for the unstructured versions, this was 90.7%. If a problem did not address a learning goal, then it was in most cases also of low complexity. The learning goals we wanted to be addressed within a context, were addressed in at least some of the problems posed for all learning goals, except one; this exception was due to our fault of not providing a good context for it.

A chi-square test was applied to analyze the differences in coverage of learning goals between structured and unstructured problem-posing exercises. From this, we conclude that there was no significant difference between the number of posed problems that covered the learning goals in structured and unstructured problem-posing exercises ($p = 0.315$).

Complexity of problems posed

To analyze the differences in complexity between structured and unstructured problem-posing exercises, we applied a Mann-Whitney U test to the coded problems. This gave $M - W = 641, p = 0.048$, which means that there is a significant difference in complexity between problems posed as a result of structured and unstructured prompts. The results in Table 3 support that problems posed in response to structured prompts are generally of higher complexity than those posed in response to unstructured prompts.

Problems posed in response to an unstructured prompt tend to be solvable by routine. For example, in the first context a student posed the problems: a. compute the expectation value; b. compute the quadratic deviations; c. compute the variance. The formulation of the structured prompt prevents posing such routine questions because the routine steps are already included as initial problems in the prompt (see Table 2).

Table 3. Relative frequencies of complexity for structured and unstructured exercises

	Complexity			Total
	Low	Moderate	High	
Structured	48,5%	36,4%	15,2%	100%
Unstructured	68,0%	30,0%	4,0%	100%

When students pose complex problems, they do this by combining the context with new elements. For example, one student posed the following problem within the second context: A box contains 20 soaps, of which 6 are blue, 7 are red and 7 are green. What is the probability of grabbing a red soap that weighs more than 125 grams two times in a row (not putting soaps back in the box)? So the student combined discrete and continuous probability.

The solvability of posed problems

We found that 69.7% of the problems posed in response to a structured prompt were solvable, whereas 94.3% of the ones from unstructured prompts were. To analyze the differences in solvability between structured and unstructured problem-posing exercises, we applied a chi-square test. There was a significant difference in solvability between problems posed resulting from structured and unstructured prompts ($p = 0.002$).

The unsolvable problems could be categorized into two categories, namely a category of problems that revealed a misconception or misunderstanding of concepts and a category of problems that were poorly formulated, but otherwise sound. An example of the first category: a student introduces the following discrete probability distribution:

X	0	1	2	3
$H(x = x)$	$1 - h$	h	$h + 1$	$h + 2$

Firstly, we see the notational issues in the first column, which should contain "x" and "$P(X = x)$". This indicates that the student does not understand the role of the random variable X and the variable x. Moreover, adding up the probabilities, one finds $2h + 4$, which should equal 1 (hence $h = -\frac{3}{2}$). However, the student means h not to be determined, and may not realize that the probabilities should add up to 1.

An example of the second category: Compute the revenue of green soaps with a probability of 0,115069670222. This is poorly formulated, and hence unsolvable, since it is not clear what the probability applies to. It could be the probability that a soap is sold to a customer, or the probability that the green soap has a certain weight, etcetera. The answer model revealed a consistent interpretation showing the student's intention, hence this problem allowed us to assess the student's progress on both learning content and mathematical problem formulation.

Conclusion and discussion

In conclusion, we state that structured prompts seem more suitable for assessing, for three main reasons. Firstly, structured prompts invite more complex problems, which in turn show more of the students' capabilities. Secondly, Structured prompts lead to significantly more unsolvable problems than unstructured prompts. This may seem bad, but it is good from an assessment point of view: those problems are usually not fundamentally unsolvable, and how the problems are unsolvable reveals misunderstandings and misconceptions of students. This may be caused by students challenging themselves more with structured prompts. This is in contrast with Mishra and Iyer (2015), who observed that, with semi-structured prompts, students would produce problems below their capabilities, as mentioned in the theoretical background. Thirdly, structured prompts lead to more problems that cover learning goals—though not significantly more. Either way, for both types of prompts the context of the task, combined with the context of the presentation of the task, namely as part of a work on a chapter on statistics and probability, ensured the problems posed revealed students' progress with respect to the learning goals of the chapter. In most cases, the problems addressed the learning goals we had envisaged, though in some cases this potential was not realized.

Our results were obtained specifically with high-achieving students enrolled in a hybrid national course. Also, the topic was specific: probability. However, we believe the conclusion on the impact of structuring the prompts holds beyond these specifics. Yet, the effect of structuring might wane after students get used to problem-posing, and know what is expected of them. From a new study, we have indications that this may happen within two or three iterations.

Since this study did not involve classroom situations, for future research we would be interested in the impact of exchange between students with respect to the problems they pose, in the line of Kwek's work (2015). Such discussion might reveal more of students' thinking and invite them to improve their problems.

Acknowledgments

The authors would like to thank Benjamin Rott for suggestions and discussions concerning this study.

References

Baumanns, L., & Rott, B. (2020). Rethinking Problem-Posing Situations: A Review. *Investigations in Mathematics Learning, 13*(2), 59–76. https://doi.org/10.1080/19477503.2020.1841501

Baumanns, L., & Rott, B. (2021). Developing a framework for characterizing problem-posing activities: a review. *Research in Mathematics Education, 24*(1), 28–50. https://doi.org/10.1080/14794802.2021.1897036

Biggs, J & Tang, C. (2011). *Teaching for Quality Learning at University*, McGraw-Hill and Open University Press, Maidenhead.

Kwek, M.L. (2015). Using Problem Posing as a Formative Assessment Tool. In: Singer, F., F. Ellerton, N., Cai, J. (Eds.) *Mathematical Problem Posing*. Research in Mathematics Education. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6258-3_13

Mishra, S., Iyer, S. (2015). An exploration of problem posing-based activities as an assessment tool and as an instructional strategy. *Research and Practice in Technology Enhanced Learning*, *10*(5). https://doi-org/10.1007/s41039-015-0006-0

Silver, E. A (1994). On mathematical problem posing. *For the Learning of Mathematics, 14*(1), 19–28.

Silver, E. A., & Cai, J. (1996). An Analysis of Arithmetic Problem Posing by Middle School Students. *Journal for Research in Mathematics Education*, *27*(5), 521-539. https://doi.org/10.2307/749846

Silver, E. A., & Cai, J. (2005). Assessing students' mathematical problem posing. *Teaching Children Mathematics*, *12*(3), 129–135. https://doi.org/10.5951/TCM.12.3.0129

Stoyanova, E., & Ellerton, N. F. (1996). A framework for research into students' problem posing in school mathematics. In P. C. Clarkson (Ed.), *Technology in mathematics education* (pp. 518– 525). Melbourne, Mathematics Education Research Group of Australasia

Stoyanova, E. (1997). *Extending and exploring students' problem-solving via problem-posing*. [Doctoral dissertation, Edith Cowen University]. Research Online Institutional Repository. https://ro.ecu.edu.au/theses/885/

Zhang, H., & Cai, J. (2021).Teaching mathematics through problem posing: insights from an analysis of teaching cases. *ZDM Mathematics Education* 53, 961–973. https://doi.org/10.1007/s11858-021-01260-3