How to STACKrate

The Development of an Intuitive Evaluation Tool

Jonas A. Lache(✉), Daniel Meißner
Ruhr-Universität Bochum, Bochum, Germany
`jonas.lache@ruhr-uni-bochum.de`

Abstract—Understanding the students' perception of STACK tasks is vital for optimizing the students' learning. Questionnaires are a common method to learn about their experiences, but they are time-consuming for teachers and students. Plus, they are usually answered with a certain delay after students have completed the task. To solve these problems, we developed STACKrate, a JavaScript-based tool that allows teachers to easily integrate a collection of evaluation questions directly into STACK questions. Thus, the evaluation is done directly after the students completed their work on the tasks. The tool uses a star rating principle and is designed to be intuitive and easy to use for students. In this paper, we describe the features and benefits of STACKrate and tell about the use of the tool in a course at Ruhr-Universität Bochum, Germany. We then discuss the added value of STACKrate in evaluation over the use of questionnaires alone.

1 Introduction

When using digital learning materials in courses, it is sensible to evaluate them to gain knowledge about the students' perceptions. And especially in the case of materials used for the first time, an evaluation can be highly valuable for finding bugs and errors. In this paper, we focus on digital mathematics tasks created with STACK. For more about the task tool STACK, see [1]. For the evaluation of STACK tasks, the most common method probably is the use of surveys which are conducted at the end of a semester, cf. [2]–[4]. While these are very sensible for learning about the general reception of the tasks as an overall package, the approach is not optimal for evaluating individual tasks or even components of tasks. One reason for this is that at the end of the semester, the completion of the tasks is often many weeks ago. However, for students to remember all relevant details, the time between processing a task and the evaluation should be as short as possible, cf. [5]. A solution could be to provide the students with weekly surveys, but this would be very time-consuming for both teachers and students. In addition, the literature shows that online surveys often have low response rates [6]. To circumvent these issues, we searched for a solution that enables

students to rate STACK tasks as soon as possible after completion. The evaluation should also be quick and easy to reach high response rates. Lastly, it should be easy to implement for task creators. These considerations led to the development of the evaluation tool STACKrate. It is a JavaScript-based tool that uses a star rating principle. In other words, the students can answer evaluation questions by selecting between one and five stars. STACKrate is distributed under the GPL-3.0 license. That means that it is available as open source and free of charge.

2   The evaluation tool STACKrate

In this section, the tool is described in detail. First, the essential design decisions and the basic features of STACKrate are explained. Then we explain how to set up an evaluation and cover some technical details.

2.1   Features of STACKrate

To align STACKrate with the above demands, the decision was made to place the evaluation questions directly into the question or feedback text of a STACK question. Hence, the students can answer the evaluation questions immediately after completing their work and remember important details about the task. To make the students' process of rating as easy and as quick as possible, the star rating principle was chosen. In the best case, only two clicks are necessary to finish the rating. Due to the low effort, higher response rates can be expected. Another benefit is that students don't need any explanations because star ratings are a popular and well-known mechanism. In a STACK question, the STACKrate evaluations appear in a highlighted box (see Figure 1).
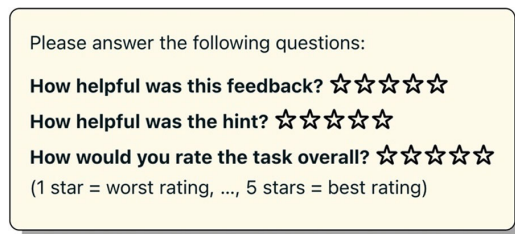

Fig. 1. Screenshot of a STACKrate evaluation box

While a STACK question can include more than one STACKrate evaluation box, each of them can again include an arbitrary number of rating questions. This makes the use very flexible. For example, different components or subtasks of a STACK question can be evaluated separately.

Beyond the possibility to choose the rating questions, there is a variety of customization options for a STACKrate evaluation. For example, it is possible to add an open text field where students can comment the STACK question or explain their rating (see Figure 2). This feature proved very valuable for our use case (see section 3).
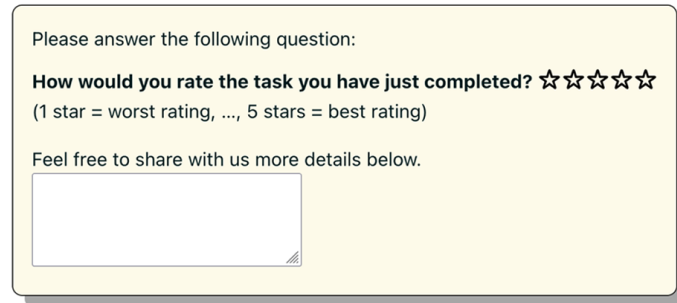
Fig. 2. Screenshot of a STACKrate evaluation box with open text field

To make STACKrate accessible to a larger group of users, we made it possible to change the language of the text snippets that STACKrate creates automatically. Currently the following languages are available: English, German, Dutch, Spanish, Portuguese, and French. Besides, the appearance of the evaluation box can be customized using CSS. In the following subsection, the implementation of a STACKrate evaluation is described and some details on technical issues are discussed.

2.2 Implementing a STACKrate evaluation

To implement a STACKrate evaluation, neither programming knowledge nor additional Moodle plug-ins are required. Task creators write their evaluation questions with special HTML markup which—together with JavaScript—takes care of the layout and functionality of the evaluation form. A hidden input field is used to store the rating results together with the other students' answers in the Moodle database. Therefore, there is no additional infrastructure necessary for storing the evaluation results. This makes it very quick to set up an evaluation form for a question.

In the following, we describe how a STACKrate evaluation can be configured. The setup consists of two steps. The first step is to add the JavaScript module and a hidden STACK input field. To do this, the following code snippet must be added to the question text:

```
<p class="ratingResults">
  [[input:ans_survey]] [[validation:ans_survey]]
</p>
<script src="https://www.rub.de/stackrate-
    maths/stackrate@0.1/stars.js">
</script>
```

The second step is more interesting: The evaluation questions are inserted into the HTML markup of the rating form. The markup is either directly placed in the question text or—if students are only allowed to rate the questions under certain conditions—to the desired feedback text. It is structured as follows:

```
<div class="stars-rating" data-name="my_id">
  <p>Please answer the following questions:</p>
  <div class="question">
    How helpful was this feedback?
    <div class="stars"></div>
  </div>
  <div class="question">
    How helpful was the hint?
    <div class="stars"></div>
  </div>
</div>
```

To correctly enter the markup into the question text, we recommend using the Moodle plain text area editor and set it to HTML format. More on the setup and additional settings can be found in the documentation of STACKrate at https://www.rub.de/stackrate-maths.

The HTML markup is designed so that certain customization is possible. As stated in the previous subsection, an arbitrary number of evaluation questions can be used. Additionally, task creators can include a text field for free comments. It can be used by adding the with-comment class to the div element which contains the whole rating form. Dealing with the different languages as described previously is easy as well. By default, the language of the Moodle site is used. If it is impossible to determine the language of the site, English is used by default. However, the language can also be set manually by adding the lang attribute to the main div element (eg., `lang="de"` for German). In addition, the appearance of the rating form can be customized using CSS.

During the rating process, the rating results are written to a hidden STACK input field as a string. This string encodes the rating results in a JSON format. When the task is submitted, the content of this hidden input field is stored together with the other answers in the Moodle database. From there it can be exported via the quiz' answer report to a CSV file. The STACKrate package contains a little Python script to extract the results of the rating from the files downloaded from Moodle. It also computes some basic statistics, like the arithmetic mean and standard deviation for the given number of stars. The script is easy to use and can usually be invoked without using the terminal. A detailed step-by-step guide on how to use it can be found in the STACKrate documentation.

3 Using STACKrate in courses: an example

STACKrate was used at Ruhr-Universität Bochum, Germany, in the winter semester of 2021/22. The goal was to evaluate STACK tasks which had been newly created for the project OER.Stochastik.nrw. In this project, we develop STACK tasks but also interactive applications and instructional videos. They are all from the field of probability theory and statistics. The project team includes mathematicians and mathematics educators from three German universities. All materials will be published as Open Educational Resources (OER) on the portal ORCA.nrw in 2023. More about the project

can be found in [7] and in section 5. In this section, however, we give some insights into the use of STACKrate in a course and the collected data. Then, we illustrate the benefits they brought for finding bugs and errors and improving the tasks.

We used STACKrate in a lecture on engineering statistics. With well over 400 students in attendance, the course was quite large. After completing a STACK task, we asked the students to give star ratings. They also got the option to write an additional comment. This way, a total of 21 STACK tasks were evaluated using STACKrate. Added together, 417 students have given at least one rating. The total number of evaluations the students gave is 6147, which makes an average of about 293 ratings for each STACK task. Each evaluation consisted of 2–3 evaluation questions. Besides, we received 716 comments from the open text fields. That makes an average of about 34 comments a STACK task. In the following, we illustrate the value of STACKrate using the example of the task "Communication Channel". This task is one of 12 tasks that include adaptive flow control as introduced in [8]. In detail, that means when the students fail to answer a problem correctly, they enter a sequence of subtasks that represent the individual steps necessary to solve the initial problem. After successfully working on the subtasks, the students are asked to answer the initial task again.

The topic of the task "Communication Channel" is the binomial distribution. First, the students are supposed to specify the appropriate distribution of the random variable described in the question text. Second, they must determine the expected value $E(X)$ and the variance $\text{Var}(X)$. Lastly, the probability $P(X \geq E(X))$ is to be specified. When the students fail answering the third subtask correctly, they are provided with two intermediate steps: specifying the probability mass function and selecting the correct formula for the probability of the counter event. Figure 3 shows the structure of the task "Communication Channel" and a screenshot.
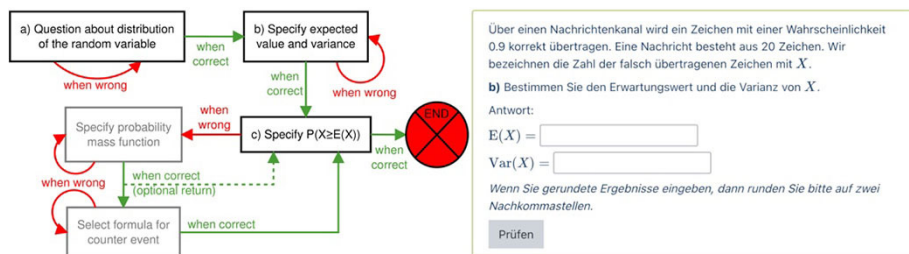

Fig. 3. Structure and screenshot of the STACK task "Communication Channel"

The STACKrate evaluation of all tasks with adaptive flow control consisted of the following two rating questions (translated from German):

1. How helpful did you find the feedback in the task to better understand the mathematical content?
2. How do you rate the task overall?

When considering only the first attempt of each person, we received n=260 ratings, making a response rate of 71%. The arithmetic mean and standard deviation of the number of stars given (reminder: between one and five stars could be selected) were

$AM_1=4.25$ ($SD_1=1.15$) for the first and $AM_2=4.24$ ($SD_2=1.14$) for the second rating question. The median was $Me=5$ for both questions. These results suggest that the task "Communication Channel" was well perceived by most of the students. Figure 4 shows the absolute frequencies of the number of stars given by the students.
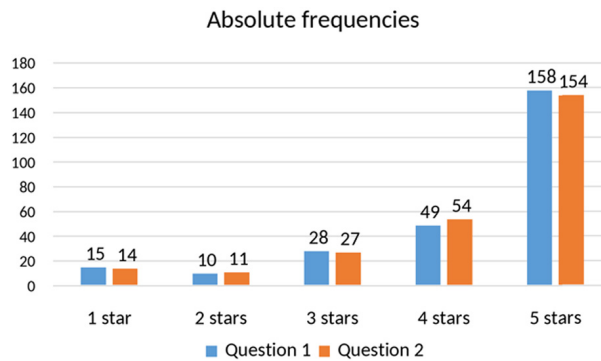
Absolute frequencies

Fig. 4. Distribution of the number of stars given for the task "Communication Channel"

In addition to the ratings, the students had the option to tell more details using an open text field. For this STACK task, we received 21 comments in which different aspects were present (all the following comments are translated from German): Some students praised the STACK task, for example, "great, thank you, keep it up" or "I am deeply impressed. Can always take place online in this form". Some students also mentioned specific aspects they liked about the task: "It was very good that we could enter the individual probabilities as a sum at the end". However, the most helpful comments turned out to be the ones in which aspects of the STACK task were criticized. For example, the comment "C was very complicated" made us aware of a possible problem with the difficulty of the third subtask. While we still do not know why the student perceived the task as too difficult, the following comment might tell us the reason: "It's a bit annoying how far the pre-programmed random variables go. When people here get $X \geq 2$ they're done in 1 min, while with $X \geq 6$ you're busy typing forever without really developing any additional understanding". A review of the STACK task following this comment revealed that the randomization of the task actually led to different levels of difficulty. Accordingly, the comment made us aware of a problem with the STACK task that previously slipped through our quality assurance. This case is an example of how valuable comments like this can be for improving the materials.

4 Conclusion and outlook

The decision to develop STACKrate was a trade-off between the time we put into STACKrate and the time that went into creating our learning materials. The effort that has gone into the development of STACKrate, however, has certainly paid off. We have got highly valuable insights into our students' perception of the tasks. Not only were we able to understand how students felt about our material—especially the individual

feedback provided in STACK questions—but student comments helped us identify problems with, for example, the randomization issue mentioned in the previous section. Finally, the direct embedding of the evaluation questions into the assignment itself yielded high response rates. Since reducing the labor to include evaluation questions was a major development goal, setting up the evaluation forms for our 21 STACK tasks went fast and reliable. Therefore, STACKrate is going to be used in other projects and in teaching at Ruhr-Universität Bochum as a supplement to conventional surveys.

If others want to use STACKrate in their own projects or teaching, this can be easily done. We have designed STACKrate so that it has a low-threshold entry and evaluation forms can be set up without any programming knowledge. An extensive online manual with screenshots and video tutorials on our website https://rub.de/stackrate-maths can help to get started. How to use the JavaScript module as well as the Python script is explained in detail. Since all necessary files are hosted by Ruhr-Universität Bochum, there is no need to setup any additional infrastructure.

We are also interested in further development of STACKrate and the adaption to new use cases. We appreciate everyone who joins us working on STACKrate. Contributing new translations, for example, is possible without any programming skills. One idea to extend STACKrate is to add the possibility to use an external database to save the results of the rating. With a feature like this, the evaluation of tasks across different Moodle tests and courses or even across Moodle sites at different universities could be simplified.

5 Acknowledgment

6 References

[1] C. Sangwin, "Computer Aided Assessment of Mathematics Using STACK", in Selected Regular Lectures from the 12th International Congress on Mathematical Education, Cham: Springer International Publishing, 2015, pp. 695–713. https://doi.org/10.1007/978-3-319-17187-6_39

[2] Ö. Genc, I. Metzler, and U. Reif, "Effects of digital tasks on learning math for first-year engineering students", in Contributions to the International Meeting of the STACK Community 2021, Tallinn, Estonia, 2021. https://doi.org/10.5281/zenodo.4916085

[3] K. Derr and M. Weigel, "University preparation with STACK: Evaluation of different question types in a self-study environment", in Contributions to the 3rd International STACK Conference 2020, Tallinn, Estonia, 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3946165

[4] R. Stetzka and E. Thevanesan, "3 Years of STACK - An experience report on the implementation and use of formative randomized online tests", in Contributions to the International Meeting of the STACK Community 2021, Tallinn, Estonia, 2021. https://doi.org/10.5281/zenodo.4916075

[5] F. Faulbaum, P. Prüfer, and M. Rexroth, Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität. Wiesbaden: VS Verlag für Sozialwissenschaften, 2009. [Online]. Available: http://link.springer.com/10.1007/978-3-531-92327-7; https://doi.org/10.1007/978-3-531-91441-1

[6] J. Daikeler, M. Bošnjak, and K. Lozar Manfreda, "Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates", J. Surv. Stat. Methodol., vol. 8, no. 3, pp. 513–539, Jun. 2020. https://doi.org/10.1093/jssam/smz008

[7] J. Lache, K. Rolka, M. Kallweit, H. Dehling, and D. Meißner, "Open educational resources for engineering statistics", in Blended Learning in Engineering Education: challenging, enlightening – and lasting? Berlin, Dec. 2021, pp. 995–1004.

[8] M. Altieri, J. Horst, M. Kallweit, K. Landenfeld, and M. Persike, "Multi-step procedures in STACK tasks with adaptive flow control", in Contributions to the 3rd International STACK Conference 2020, 2020. https://doi.org/10.5281/zenodo.3944786

7 Authors

Jonas A. Lache is a research assistant at Ruhr-Universität Bochum, Faculty of Mathematics, Postfach 70, 44780 Bochum, Germany. He is currently a doctoral student in the field of mathematics education on digital mathematics tasks (email: Jonas.Lache@ruhr-uni-bochum.de).

Dr. Daniel Meißner is a postdoctoral researcher at Ruhr University Bochum, Faculty of Mathematics, 44780 Bochum, Germany. He works in the field of probability theory and its applications and the development of digital learning materials like STACK questions (email: Daniel.Meissner-i4k@ruhr-uni-bochum.de).