Evaluating Engagement and Learning Based on a Student Categorization using STACK, Exam Data, Key Informant Interviews, and Focus Group Discussions

https://doi.org/10.3991/ijet.v17i23.36619

Zevick Otieno Juma[1(✉)], Mildred Atieno Ayere[1], Michael Obiero Oyengo[1], Georg Osang[2]
[1]Maseno University, Maseno, Kenya
[2]IDEMS International Company, UK
jumazevick@gmail.com

Abstract—Systems for Teaching and Assessment using Computer Algebra Kernel (STACK) is a computer-aided assessment plug-in for the Moodle learning management system that provides sophisticated tools for student assessment in mathematics and related disciplines, with emphasis on formative assessment. In the last four years, IDEMS international has supported the School of Mathematics at Maseno to integrate STACK and use it in the teaching, learning, and assessment of undergraduate students in nine courses. One of the courses was "Introduction to Complex Analysis", a third-year course shared by students taking mathematics-related programs from different faculties within Maseno. This paper reports on an evaluation of learner behavior in the Complex Analysis course using data from the STACK weekly quizzes done in that course, the final exam, 20 key informant interviews, and 4 focus group discussions.

Keywords—using Moodle STACK at Maseno, learner engagement with STACK, online assessment in STEM

1 Introduction

The use of digital technology in teaching, learning, and assessment is becoming more common today [2], [6], [9]. This is due to a number of advantages that come with it such as; boosting collaboration, preparing students for the future, and providing a more engaging learning environment among others. STACK is a computer-aided assessment plug-in for the Moodle and ILIAS learner management systems. It adds a sophisticated assessment in mathematics and related disciplines, with a lot of emphasis on formative assessment underpinned by computer algebra. Figure 1 shows an example question, along with the student's response and feedback. Notice how in this example learners must enter their solution as a mathematical expression rather than selecting an option like in a multiple-choice question.

Using the open-source computer algebra system "maxima", STACK allows question authors to create randomly generated mathematical questions within structured

templates, check student answers submitted within that question, and provide immediate feedback taking into account the mistakes made.

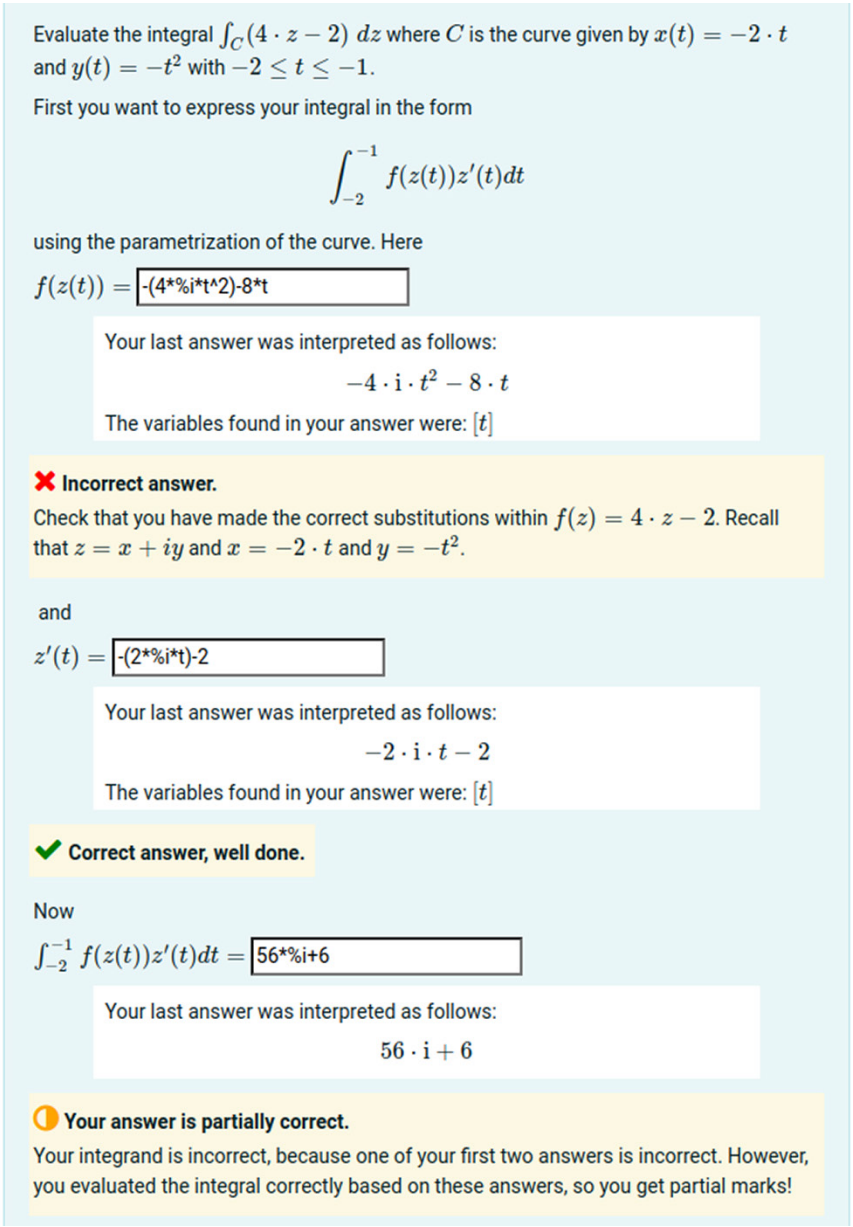


Fig. 1. An example STACK question with corresponding feedback

Research on STACK has been widely published by authors from various contexts, showing its use in various education settings all over the world [1]–[9]. For the last four years, Maseno university has used STACK in formative assessment in nine

undergraduate mathematics courses. The long-term goal is to integrate STACK in the teaching, learning, and assessment of students at Maseno. It is thus important to study the effectiveness of STACK quizzes as resources to student learning. Our hypothesis was that students' engagement with STACK quizzes influences their learning and performance in the end-of-course exams. In 2021, we investigated student engagement with STACK resources in one of the courses "Introduction to Complex Analysis", taught to third-year students at Maseno University. Teaching was primarily done face-to-face and continuous assessment was carried out digitally using STACK quizzes. The subsequent section of this paper describes the Maseno context and explains why STACK is valuable in teaching and assessment, therefore, evaluation of its usage becoming necessary.

2 The context

Maseno University is a public learning institution founded in 1991 and located in Kisumu County, Kenya. It has two main branches, the Kisumu Campus, located in Kisumu city, and the Main Campus, found in Maseno town. Every year more than 5500 students are enrolled in various degree programs at Maseno.

The school of mathematics statistics and actuarial science (SMSAS), offers six undergraduate programs, four Masters degree programs, and three Ph.D. programs. It has sixteen teaching staff and four part-time lecturers. The school provides services to other faculties within Maseno whose students take mathematics-related degree programs and thus, are required to take compulsory mathematics courses as part of their study requirements. As a result, these classes tend to be quite large with 500 to 1,200 students in a single class. SMSAS lecturers are required to teach at least four courses in a single semester with no support. Thus, students do not have the chance to engage meaningfully through continuous formative assessment with timely feedback from such assessments, if any. Admittedly, instructors may administer one or two continuous assessment tests (CATs) per course in such large classes giving students very limited opportunity to practice and enhance mastery of basic-to-advanced concepts in mathematics.

Therefore, lecturers are tasked to seek creative ways to ensure the delivery of content and conduct formative assessments in their courses. For the last four years, SMSAS has used STACK in formative assessment in nine undergraduate courses. The long-term goal is to integrate STACK in the teaching and assessment of undergraduate mathematics courses, especially the common courses where the student population is 700 per class, on average. Some of the expected challenges in such environments are access to devices, the internet, and a lack of clear objectives within the student organization on how useful STACK could be in formative assessment. The methodology section explains the various steps followed while conducting this research.

3 Methodology

For a period of 8 weeks, students took two quizzes each week, Mastery and Test Quizzes, deployed through STACK via the Maseno Moodle site. Exam revision

quizzes, which did not follow the Mastery-Test Quiz format, were also provided to the students on the 9th week to help them prepare for the final exam. Mastery quizzes were set such that students had unlimited number of attempts on them, and the maximum grade achieved recorded. They were made available the entire semester. Students, on the other hand, had to score at least 70% on that week's mastery quiz in order to access the test quiz for that same week. Test quizzes could only be taken once and were only available on weekly basis, after which they were closed at the end of the week. By design, most questions on the test quiz were similar to the ones on the mastery quiz. This was done to encourage students to gain confidence on the mastery quizzes before attempting the test quizzes. STACK quizzes contributed 30% of the final course grade, with both the mastery and test quizzes contributing 15% each.

At the end of the semester, students sat for an exam, which was a traditional paper-based written exam made up of questions mostly from the weekly STACK quizzes. The final exam contributed to 70% of the final course grade, as is the Maseno University policy on the examination of undergraduate students.

Student behavior on the weekly STACK quizzes was recorded in the Maseno moodle Learning Management System. Of particular interest to this study was the frequency with which students engaged with the quizzes, the duration taken to complete each quiz, and the score on each of the quizzes. An analysis was done on the findings with the aim of drawing a correlation between student behavior on the weekly STACK quizzes, and the end-of-semester exam.

There were instances in the correlation analysis where student results defied the trend. We requested 24 interviews to investigate this, and 20 students responded and attended the interviews. We went ahead and asked 36 students to volunteer in one of the four Focus Group Discussions having 9 participants spread across each group (1 male, 1 female, and 2 mixed-gender groups), where more qualitative data was collected from the participants, in an attempt to corroborate the responses given by students in the interviews. 32 students turned up for the FGDs. In both the interviews and the 4 FGDs, participatory research design was used to encourage students to help provide insights into different behavior patterns observed in the course. The quantitative data were analyzed first, and the results are presented in Section 4.1. A qualitative analysis was also carried out, and all student responses have been presented thematically in Section 4.2.

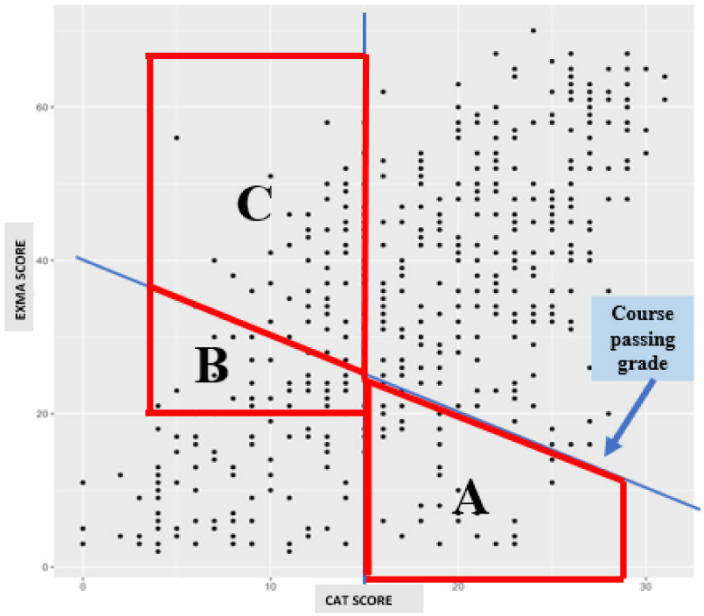
4 Results and discussions

The first half of this section presents a quantitative analysis of students' results, looking at scores, the number of attempts, and attempt durations of the STACK quizzes, as well as the exam scores. As mentioned earlier, we also observed groups of students defying those trends, warranting closer investigation, this has been outlined in the second half of this section.

4.1 Quantitative analysis

In this study, a comparison was made between students' final exam results to their STACK quiz results, the number of attempts on the STACK quizzes, as well as total time spent on the STACK quizzes.

Figure 2 shows the relationship between student performance in the weekly STACK Quizzes computed out of 30 marks and the traditional paper-based written exam done at the end of the semester (out of 70 marks). The overall pattern on the scatter plot shows a positive correlation between the two test results. We also observed 3 groups of students who defied the trends, see Figure 2 with labels A, B and C. The dots shown above the diagonal line labeled "Course passing grade" represents students who attained the course passing grade, which was 40%. The converse applies to those below the diagonal line.



A – Students who scored above average on the STACK Quizzes and failed the course.

B – Students who scored above 20 marks in the exam and had below-average test scores in the STACK Quizzes, thus they still failed the course.

C – Students who had below-average test scores in the STACK Quizzes, and managed to pass the course because they scored sufficiently well on the exam

Fig. 2. Categorization of students using their scores in STACK Quizzes and the Final Exam

Figure 3 shows the relationship between the overall frequency of engagement with the weekly Mastery Quizzes done the entire semester and the corresponding performance in the exam. From the scatterplot, students who engaged more with STACK performed better on the final exam. Also, we closely investigated a group of students who had a relatively higher number of quiz attempts but with low exam scores (labeled D in Figure 3). Our hypothesis was that they didn't seem to learn from the feedback in the quizzes. We conducted a qualitative analysis to find out why.

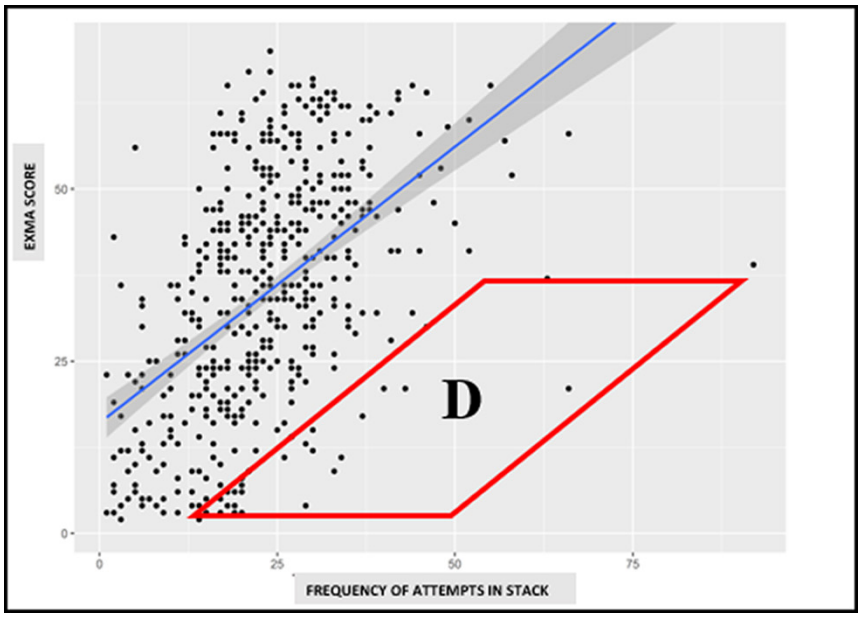


Fig. 3. Student categorization using the frequency of attempts in STACK and exam data

In addition to the total number of attempts, Moodle records the duration of each quiz attempt. Moodle saves students' answers and allows them to resume later, but only records the duration between opening and submitting a quiz. We observed that many students worked on the quizzes for multiple days. Thus, the data was not indicative of the time spent actively working on the quiz, making it difficult to use in a meaningful way.

4.2 Qualitative analysis

We conducted interviews together with FGDs to gain a better understanding of how students are using STACK, and to find possible explanations concerning the findings in Figures 2 and 3. The findings have been grouped under four main thematic areas titled as follows: Criticisms of STACK, Challenges of using STACK, Value of STACK, and Potential of STACK. For each topic, we discuss the students' responses and present a few quotes that are representative or of particular interest.

***a*) Criticisms of STACK**

According to 17/20 and 31/32 of the responses in the interviews and FGDs respectively, the majority of students who did well on the STACK quizzes and failed the course cheated, see Figure 2 (group A). 5 out of the 17 who responded in the interviews added that they were aware of fellow classmates who "paid" their colleagues to take the quizzes for them. The aforementioned 31 participants in the FGDs agreed that it was not unusual for students to pay their colleagues to do their homework and they didn't seem to resent that idea.

To investigate this behavior, we had a closer look at students who scored more than 15/30 marks on the CAT, and within this group, compared those who performed badly in the exam (≤15/70 marks) to the average. While the average CAT score of those students who performed poorly on the exam (20.6/30) was comparable to the average in the entire group (21.0/30), we observed that most of the time, they had only one recorded attempt per mastery quiz (median 10 mastery attempts in total, or 1.25 per quiz), while the average student in this group had multiple attempts (median 18 mastery attempts in total, or 2.25 per quiz). This supports the hypothesis of cheating coming out of the interviews: Students having the quiz done for them ("cheaters") only have one recorded attempt, as the students doing the quizzes on behalf of the cheaters are already proficient at them, however, the cheaters also perform poorly on the exam as they do not learn from the quizzes.

Some students, although just a few (about 20%) in both the interviews and FGDs, mentioned that there were some questions where feedback didn't seem to show all the computational steps leading to the final answer in the solution. This, to some extent, contributed to them having multiple attempts on the quizzes in an attempt to understand the content from the various variants of the question in the Mastery Quiz (see Figure 3). This is not a limitation of STACK per se, but just a criticism about how some of the questions were authored.

The remaining proportion who didn't seem to agree (3/20 in the interviews and 1/32 in the FGDs) cited other possible reasons as to why one could do well in the STACK quizzes and still fail the exam. They mentioned exam anxiety and lack of preparation for the exam as possible causes. Figure 4 presents a few quotes from the students' responses on this topic.

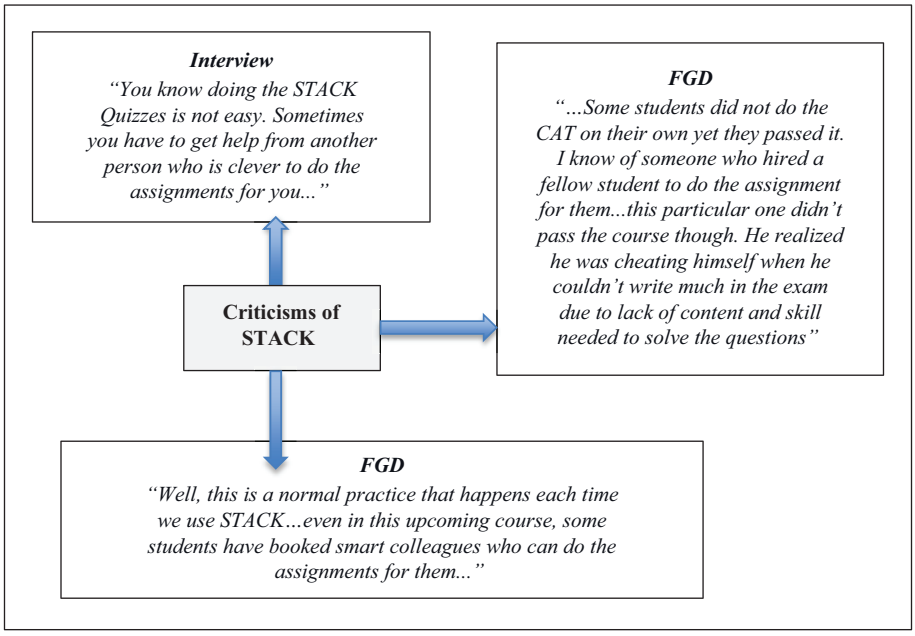


Fig. 4. Responses from students on "Criticisms of STACK"

***b*) Challenges of using STACK**

According to 15/20 of respondents in the interview, lack of access to devices and internet connectivity were one of the major contributors to why some students didn't engage more with the quizzes (see Figure 2 student categorization B and C). The feedback from the interviews were corroborated by all the responses in the FGDs, though, to a small degree, they argued that only a few students faced this challenge. They estimated that about 20% of the students had challenges with access. From all the responses, the main device used to access STACK were smartphones with students having to purchase data bundles, due to the unreliable university Wi-Fi network. The remaining 5 responses in the interviews cited low motivation, as well as disinterest in mathematics, as possible reasons why some students did not engage with the STACK quizzes. Figure 5 shows quoted responses from students on the topic 'Challenges of using STACK'.

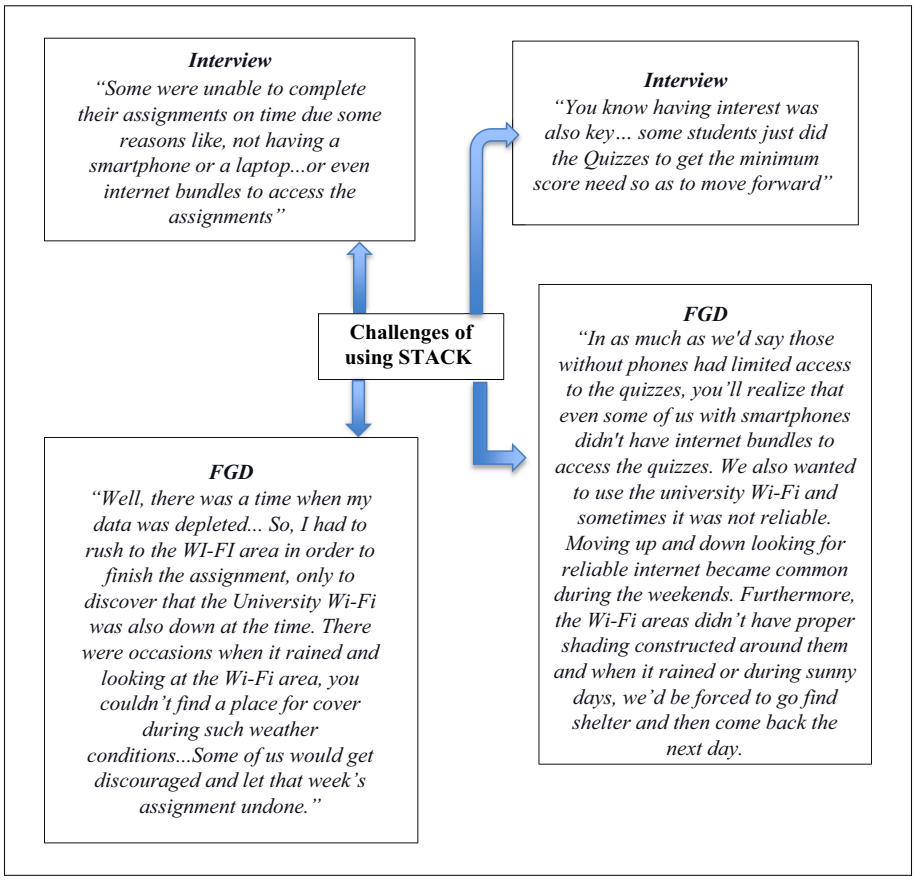


Fig. 5. Summary of participant responses on challenges of using STACK

***c*) Value of STACK**

One key aspect of STACK which was an asset to all the students in the course was its ability to provide immediate feedback and randomize questions for practice. This was

according to 15/20 of responses from the interviews and 31/32 of those in the FGDs. According to those 15/20 in the interview, most students who didn't do well on the Continuous Assessment used Mastery Quizzes (because of getting immediate feedback) for practice and preparation for the final exam. Figure 6 shows quoted responses from students during the interviews and also in the FGDs.

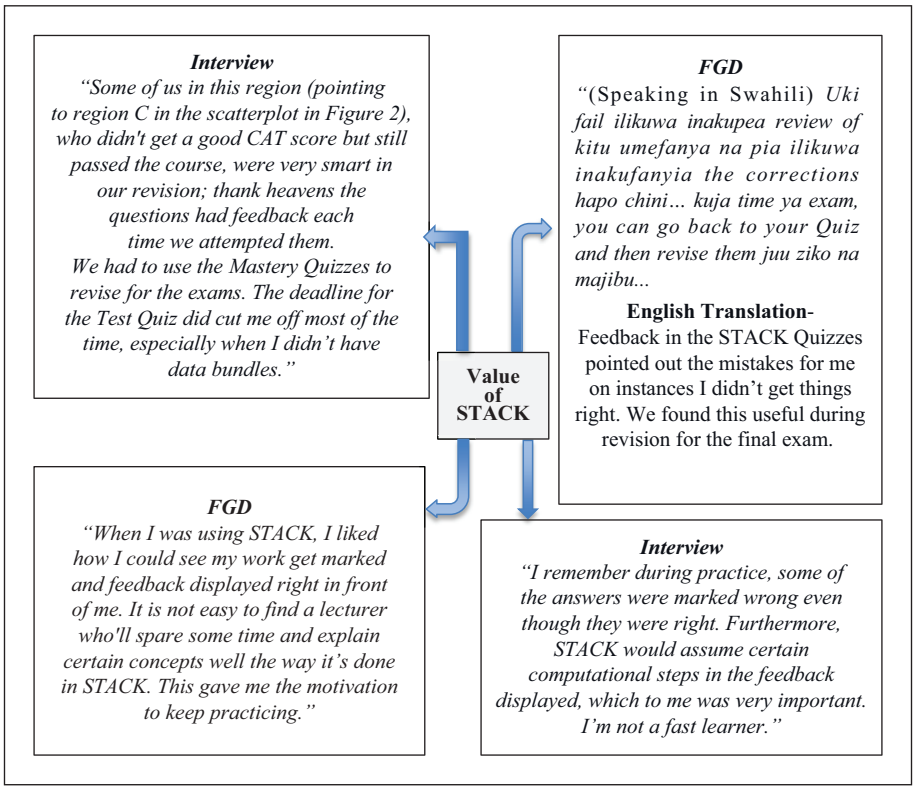


Fig. 6. Students' responses on the value of STACK

***d*) Potential of STACK**

In an attempt to find out about learner perception on the use of STACK in formative assessment, 19/20 of the respondents in the interviews recommended continuous assessment to be conducted using weekly STACK quizzes. The 20th person was not against it, but rather had a concern, which was access to devices since some students were not able to do the quizzes on time. From the FGDs, 30/32 of the students recommended that STACK should be integrated in all mathematics courses due to its ability to give immediate feedback, pointing out the mistakes, therefore, guiding the learning process, and allowing them to practice their math skills. It came out from both the interviews and the 4 FGDs, that students found using STACK rather motivating, citing that it kept them "on toes with learning and practicing mathematics". These responses suggest that, this could be a possible explanation why despite having challenges using STACK, some students managed to look beyond and see its potential in helping them

learn mathematics. Figure 7 shows some of the quotes from the student feedback on this topic.

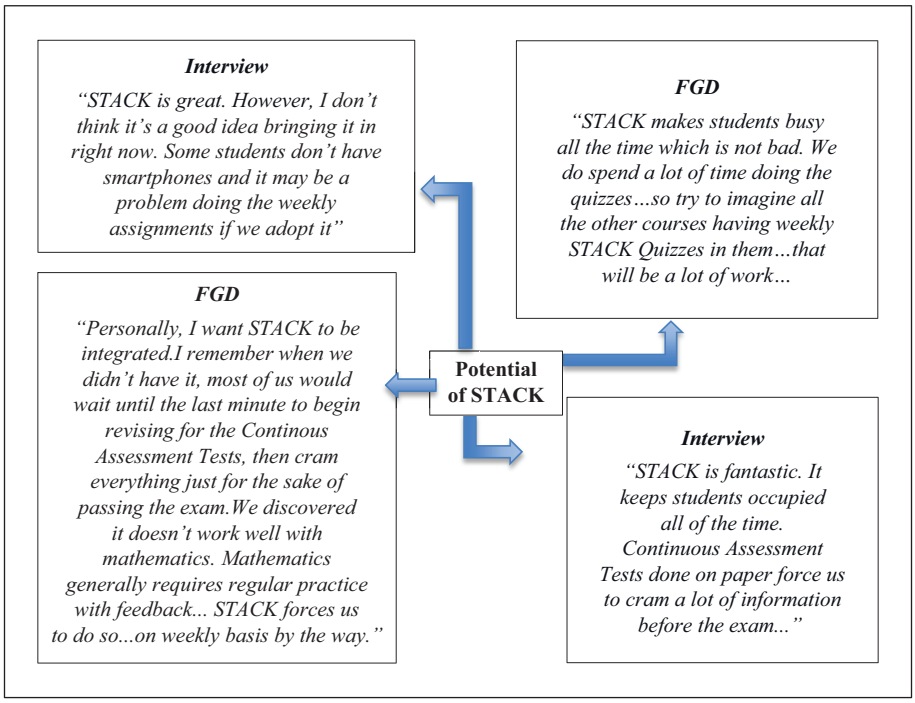


Fig. 7. Students' responses on the topic "Potential of STACK"

5 Conclusion

The data presented in this research suggests that, while some students appear to show engagement with STACK according to the Moodle report statistics on learner behavior, in the Maseno context, a notable proportion seem to have limited engagement. To some extent, it appears that a lack of engagement with STACK can be linked to issues around access, network connection, or even poor feedback.

Cheating is a vice that has come out from the findings in this study, and instructors using STACK or planning to adopt it in learning institutions with a similar context as Maseno should be aware of this. Based on the outcome concerning the issue of cheating, it is notable that students engaged in this vice are mainly cheating themselves. Creating awareness of that might discourage the behavior.

Students seemed to appreciate STACK's ability to provide immediate feedback, randomize questions for practice, and it being readily available, making it preferable to non-electronic formative assessment.

According to the student responses, despite there being challenges in the use of STACK, learners seem to appreciate and point out its potential in the Maseno context.

STACK is playing a significant role in providing individualized feedback to students, which was previously infeasible. As a result, Maseno wishes to continue its collaboration with IDEMS International, in the development of high-quality resources, while also addressing some of the issues raised in this study to continue using STACK in providing high-quality mathematics education.

6 References

[1] Bach, Stephan. (2020). *Using multiple choice questions in STACK – reasons and examples*. https://doi.org/10.5281/ZENODO.3945809

[2] Derr, K. (2019). Optimising Self-study With STACK. *STACK Online Assessment. A Collection of Case Studies*. https://www.academia.edu/44339885/optes_Optimising_Self_study_With_STACK

[3] Nakamura, Y., Ohmata, Y., & Nakahara, T. (2012). Development of a question-authoring tool for math e-learning system stack. *IADIS International Conference E-Learning*, 435–440.

[4] Oyengo, M. O., Parsons, D., Stern, D., & Sangwin, C. (2021, June 9). *Providing student feedback through electronic assessment for linear algebra at Maseno University, Kenya*. International Meeting of the STACK Community 2021. Zenodo. https://doi.org/10.5281/zenodo.5035969

[5] Sangwin, C. J. (2007). Assessing elementary algebra with STACK. *International Journal of Mathematical Education in Science and Technology*, *38*(8), 987–1002. https://doi.org/10.1080/00207390601002906

[6] Sangwin, C. J. (2010). Who uses STACK? A report on the use of the STACK CAA system. *The University of Birmingham*.

[7] Singh, M. N. (2021). Inroad of digital technology in education: Age of digital classroom. *Higher Education for the Future*, *8*(1), 20–30. https://doi.org/10.1177/2347631120980272

[8] Tomilenko, V. A., & Lazareva, E. G. (2020). Effective use of a computer grading system for teaching mathematics at a university. *ITM Web of Conferences*, *35*, 03017. https://doi.org/10.1051/itmconf/20203503017

[9] Ustinova, I., Tomilenko, V., Imas, O., Beliauskene, E., & Yanuschik, O. (2020). STACK assessment in mathematics classroom: advantages and disadvantages. *International Conference on Interactive Collaborative Learning*, 174–182. https://doi.org/10.1007/978-3-030-68198-2_15

[10] Wells, J. G. (2019). *STEM Education: The Potential of Technology Education*. Council on Technology and Engineering Teacher Education. https://vtechworks.lib.vt.edu/handle/10919/93963

7 Authors

Zevick Otieno Juma is a Masters of Education technology student at Maseno University studying on the use of technology (STACK) in the teaching and assessment of Mathematics. He is working as an intern with IDEMS International on developing STACK resources to help with teaching and learning of STEM subjects in low-resourced settings in Africa.

Mildred Atieno Ayere has a PhD in Education Technology. She is a lecturer in the Department of Educational Technology and Curriculum Studies, Maseno University.

Her primary research interests include online learning methodologies, ICT integration in pedagogy and interactive learner-centered methodologies in teaching and learning in general and Mathematics in particular. She is also the current director eCampus, Maseno University.

Michael Obiero Oyengo has a PhD in Mathematics and is a lecturer at the School of Mathematics, Statistics, and Actuarial Science, Maseno University. He is also actively engaged in STEM research, specifically mathematics education projects in low-resource settings.

Georg Osang obtained his PhD from the Institute of Science and Technology Austria, and now works as a mathematical scientist for IDEMS International. His research interests lie at the intersection of mathematics, computer science, and education.

Article submitted 2022-10-04. Resubmitted 2022-11-03. Final acceptance 2022-11-05. Final version published as submitted by the authors.