Effects of Technological Interventions for Self-regulation: A Control Experiment in Learnersourcing

Hatim Lahza
h.lahza@uq.edu.au
The University of Queensland
St Lucia, QLD, Australia
Umm Al-Qura University
Makkah, Makkah al-Mukarramah, KSA

Hassan Khosravi
h.khosravi@uq.edu.au
The University of Queensland
St Lucia, QLD, Australia

Gianluca Demartini
g.demartini@uq.edu.au
The University of Queensland
St Lucia, QLD, Australia

Dragan Gasevic
dragan.gasevic@monash.edu
Monash University
Clayton, Vic, Australia

ABSTRACT

The benefits of incorporating scaffolds that promote strategies of self-regulated learning (SRL) to help student learning are widely studied and recognised in the literature. However, the best methods for incorporating them in educational technologies and empirical evidence about which scaffolds are most beneficial to students are still emerging. In this paper, we report our findings from conducting an in-the-field controlled experiment with 797 post-secondary students to evaluate the impact of incorporating scaffolds for promoting SRL strategies in the context of assisting students in creating novel content, also known as learnersourcing. The experiment had five conditions, including a control group that had access to none of the scaffolding strategies for creating content, three groups each having access to one of the scaffolding strategies (planning, externally-facilitated monitoring and self-assessing) and a group with access to all of the aforementioned scaffolds. The results revealed that the addition of the scaffolds for SRL strategies increased the complexity and effort required for creating content, were not positively assessed by learners and led to slight improvements in the quality of the generated content. We discuss the implications of our findings for incorporating SRL strategies in educational technologies.

CCS CONCEPTS

• Applied computing → Computer-assisted instruction; **• Human-centered computing → Empirical studies in collaborative and social computing**; **Collaborative content creation**; **Collaborative and social computing systems and tools**;

KEYWORDS

self-regulation, learnersourcing, software-based scaffolding, metacognition

ACM Reference Format:
Hatim Lahza, Hassan Khosravi, Gianluca Demartini, and Dragan Gasevic. 2022. Effects of Technological Interventions for Self-regulation: A Control Experiment in Learnersourcing. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22), March 21–25, 2022, Online, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3506860.3506911

1 INTRODUCTION

Self-regulated learning (SRL) is considered an essential competence and a lifelong learning skill [36]. SRL views a learner as an agent [31] who can become proactive [37] to choose, adapt and develop strategies and instructions they need to accomplish their learning goals [31, 36]. Educators can support learners' SRL development through the use of various scaffolds to enable students to better metacognitively monitor and control their learning [4]. However, as class sizes increase, it becomes more challenging for instructors to effectively incorporate SRL strategies in class, and it becomes more difficult to measure the effectiveness of these strategies in real-life contexts (i.e., to what extent students are utilising the scaffolding artifact).

Recent advances in educational technologies enable education system designers to develop customised learning environments in which educators can readily manage the provision of SRL and metacognitive support, ranging from fixed scaffolds to dynamic adaptive scaffolds [4, 8, 23]. However, while theory about the role of SRL in enhancing learning is coherent [26], the best methods for incorporating them in educational technologies and empirical evidence of their effectiveness are still emerging [8][3, p.14]. This paper contributes to filling this research gap by empirically examining the impact and challenges of incorporating technological interventions to promote effective SRL strategies to support students in creating novel content. Implementing SRL in different learning contexts might shed light on how SRL can improve learners' performance and learning, and whether the implementation principles should differ from one context to another [14].

Engaging students in content creation, also referred to as learnersourcing [19, 22], has been demonstrated to have many benefits,

LAK22, March 21–25, 2022, Online, USA

ACM ISBN 978-1-4503-9573-1/22/03…$15.00
https://doi.org/10.1145/3506860.3506911

including development of large repositories of learning content that can be used for studying [24] or within adaptive educational systems [21]. However, in terms of quality, the contents created by students vary from those that meet rigorous standards to those that are incorrect or ineffective. We posit that part of the issue might be explained by the fact that learnersourcing systems have been offering expert-in-training students a similar interface to those offered to expert instructors. We, therefore, investigate whether designing learnersourcing interfaces that benefit from well-established SRL strategies may help students in creating content. We limited the scope of our experiment to helping students create multiple-choice questions (MCQs) (and their variants, i.e., multiple answers) and offered three strategies from the SRL literature: *planning* to help students set goals and better prepare for creating MCQs [17, 34]; *externally-facilitated monitoring*, providing students with checklists based on best practice for creating MCQs [12] to help them monitor their performance [5, 13]; and *self-assessing* to encourage students to self-assess their work and write a reflection about their performance and the process they followed to come up with the question [27, 29].

To evaluate our design, we conducted an in-the-field randomised control experiment of five conditions with 797 participants who had created 1,907 MCQs in two university-level courses. The five conditions included a control group that had access to none of the strategies for creating content, three groups each having access to one of the strategies (planning, monitoring and self-assessing) and a group with access to all of the aforementioned strategies. We explored the effect of the strategies on the process of creating content, students' perception of their benefits and their actual impact on the quality of the created resources.

2 METHOD

In this study, we explore the effect of incorporating the explicit SRL strategies of planning, externally-facilitated monitoring and self-assessing on learnersourcing. Our study is guided by the following research questions:

RQ1. How do the proposed interventions change the process and effort required for creating a learning resource?

RQ2. How beneficial do students find the proposed interventions?

RQ3. What is the impact of the proposed interventions on the quality of the resources created by students?

2.1 The RiPPLE System

We made use of RiPPLE [20] for this study. At its core, RiPPLE is an adaptive educational system, where the learning content is developed via learnersourcing. RiPPLE offers three types of learning activities, namely create, evaluate and practice. Here is how it works: students in a course create study resources such as MCQs. Figure 1-a shows the current interface for creating an MCQ without the elements inside the dashed boxes. Created resources are peer-reviewed by multiple other students. Reviewers rate the quality of the resource and their confidence in their rating and provide feedback on how the resource can be improved. RiPPLE utilises the submitted peer reviews and the algorithms discussed in [1] to make a decision about the quality of the resource. Effective resources are approved and are added to the RiPPLE repository while ineffective resources are rejected and sent back to the author for resubmission. Instructors can oversee this process using RiPPLE's AI spot-checking feature [10] to identify wrong decisions. The repository of the approved resources is then used to provide various practice opportunities (e.g., formative quizzes).

2.2 Incorporated Interventions for SRL Strategies

This section describes the design choices and techniques used to complement the current content creation interface. Three main strategies, "planning then creating", "externally-facilitated monitoring" and "self-assessing after creating", are employed in our approach. An additional interface was created to have all the components of these strategies in one place. Figure 1 illustrates the main interfaces of the addition of these strategies.

Planning then creating. The planner shown in Figure 1-b was added as a preceding step to the creation component. The first question of the planner, which focuses on the sources of inspiration for their question, was added in response to the feedback of some students who were unsure of how to create novel content. The drop-down list included options such as course textbook, lecture slides, previous exams, other students' questions and the course discussion forum to provide some examples and prompt students to review sources that could help them in creating a resource. The other elements of the planner were in accordance with best practices for creating MCQs (e.g., determining Bloom's taxonomy [18], using MCQs multiple format [2] and identifying concept and misconception [12]). For students who had a clear vision of the resource they wanted to create, the planner aimed to trigger their metacognition to help them organise their learning according to certain characteristics of MCQs (e.g., the learning level the question assesses) and make their plans explicit and available for self-monitoring and revision. For students who did not have a clear idea for creating a resource, the planner aimed to support their goal setting and planning. The effect of planning on SRL has been acknowledged in past studies [17, 34].

Externally-facilitated monitoring (monitoring). The self-monitoring tick boxes shown by the dashed boxes in Figure 1-a were added to help students leverage best practices (MCQs principles) for creating MCQs (e.g., [12]) and better monitor their performance, which may lead to the development of high-quality resources. The effect of externally-facilitated monitoring on SRL has been acknowledged in past studies [5, 13, 15].

Self-assessing after creating. The self-assessing interface that is shown in Figure 1-c was added to prompt students to review and revise their resources, which may lead to the development of high-quality resources. The interface had two parts. In the first part, students assessed their work using a rubric with a set of criteria, which was similar to those used by their peers for evaluating their work. In the second part, after self-assessing via the rubric, students would reflect on their performance and were asked how they could improve the resource as recommended by [23]. The effect of self-assessment on SRL has been acknowledged in past studies [27, 29].

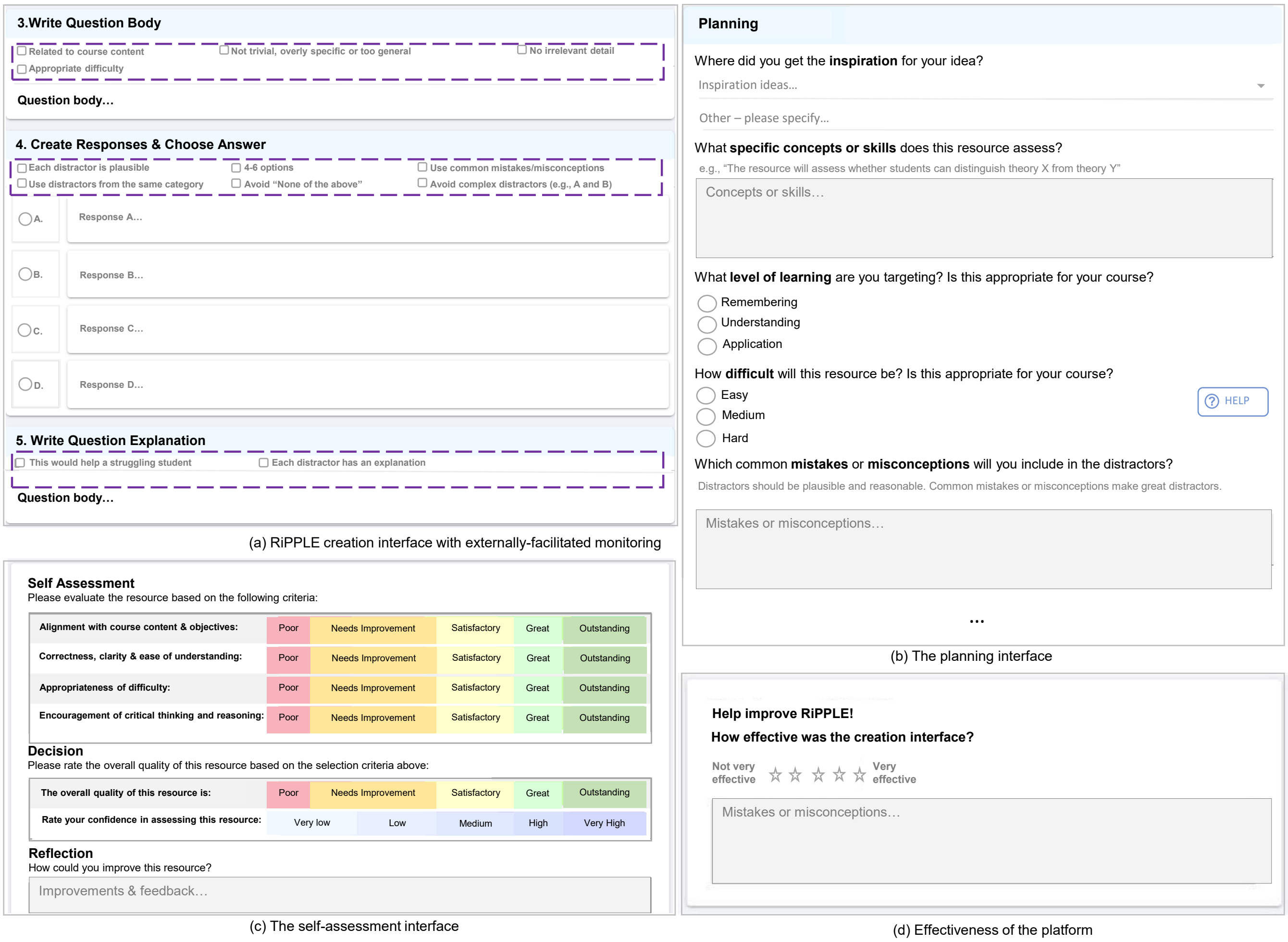

Figure 1: The main components used to create the five conditions.

2.3 Experimental Design

2.3.1 Experimental Settings. To evaluate the effectiveness of the three strategies, we conducted an in-the-field randomised control experiment of five conditions, including a control group that had access to none of the strategies for creating content (condition A); three groups each having access to one of the strategies, planning then creating (condition B), monitoring while creating (condition C) and self-assessing after creating (condition D); and a group with access to all three aforementioned strategies (condition E).

2.3.2 Research Context & Participants. To conduct the experiment, we recruited participants from two different undergraduate courses at The University of Queensland[1]. Data from the 797 students who had provided their consent in RiPPLE was used in the analysis and the reported results. Both courses used RiPPLE as an assessment tool over multiple rounds in which students' engagement with RiPPLE had a 10% contribution towards their final grade. The grade associated with RiPPLE in both courses was conditional on students' engagement with creating resources that were approved (i.e., one or more resources weigh one mark in each round) and peer evaluating the created resources. Table 1 provides an overview of the experimental groups in terms of the number of students (# students), percentage of students who at least had one full engagement with the strategies, which was used in answering RQ3 (% engaged), and number of resources (# resources) from each of the offerings. The average resources created by students were: INFS course: 2.13 ± 1.08, NEUR course: 2.57 ± 1.27 and both: 2.39 ± 1.22. We considered students who had utilised all the components of the intervention in at least one creation session as "engaged". For text fields with an open response, answers with at least three words were considered as meeting the requirement for being engaged.

2.3.3 Metrics and Analysis. In the interest of space, as the patterns of behaviour were similar across the two courses, we only report combined results based on data collected from both courses. We used the Mann-Whitney U test to perform statistical analyses of the reported results of each of the treatment groups against the control

[1]Approval from our Human Research Ethics Committee #2018000125 was received for conducting this study.

Table 1: Overview of the experimental groups

Group	INFS			NEUR			Both		
	# students	% engaged	# resources	# students	% engaged	# resources	# students	% engaged	# resources
Control	60	100	134	113	100	299	173	100	433
Planning	63	60	141	71	51	181	134	55	322
Scaffolding	63	35	119	103	46	278	166	35	397
Self-assessment	69	96	153	109	97	257	178	97	410
Combined	74	20	155	72	17	190	146	18	345
Total	329	62	702	462	66	1205	797	61	1907

group. We use $p < 0.05$ as the criterion for assessing statistical significance.

RQ1. Impact on the Process and Effort. To address RQ1, we aimed to detect the process patterns of using the interfaces and the time taken to create MCQs in the five groups. Process mining (PM) is a common learning analytics technique used to accomplish this task [32]. Subsequently, we used the timestamped interaction data in the log file to construct a First Order Markov Model (FOMM) for each of the courses and their combined data. The types of activities in the log file depended on the group condition. For instance, condition A only had one activity type, the creation, while condition B had two activities, creation and planning. To conduct the analysis, we used the pMineR package [16], descriptive analysis and Mann-Whitney U test.

RQ2. Impact on Student Perceptions. To answer RQ2, we included an optional component across all conditions that asked students to rate the effectiveness of the interface (i.e., 5-point scale) and provide written feedback before each submission (see Figure 1-d). We first used descriptive analysis, Mann-Whitney U test to examine the difference between each treatment group and the control group. We then qualitatively analysed the written feedback to elicit insights using a general inductive approach [33]. In the qualitative analysis, the authors first reviewed the 137 responses provided by 103 students to obtain an overall impression of the themes. Second, the first author identified the themes in the data through several iterations, and accordingly, we used three high-level types of codes: "technology", "pedagogy" and "generic". Finally, we utilised lexical sentiment analysis using the Sentimentr package [30] to calculate sentiment scores for the responses. What distinguishes Sentimentr from most of the existing dictionary lookup methods is that it takes into consideration valence shifters (e.g., "not" and "hardly").

RQ3. Impact on Performance. As mentioned in Section 2.1, RiPPLE utilises the submitted peer reviews and the algorithm discussed in [1] to decide about the quality of the resource. We use these inferred quality ratings to compare the quality of resources created in the control group against the quality of resources in each treatment group. Besides using the Mann-Whitney U test for statistical significance, we use Pearson's r to measure the effect size of each of the interventions. We conducted the analysis for "all students" and then separately for the "engaged" students, as explained in Section 2.3.2. We conducted the analysis separately for the engaged students as the benefits of the interventions are less likely to have had an impact on the work of students who had not explicitly engaged with them.

3 RESULTS

RQ1: Impact on the Process and Effort. Figure 2 visualises the process of how students engaged with the interventions. The visualisation of the process consists of connected oval shapes representing the type of activity and arrows showing the direction of the connection. The strength of the connection is indicated by the thickness of the arrows and the transition probability next to it. Inside the oval shapes, we added the fractional time the students spent on the activities. In most of the creation tasks in condition B, students began with the planning activity (71%), then created a question (92%) and then submitted their resources. Also, as evidence of engaging the students with planning, there were strong interactions between the planning and creating components (Pl –> Cr: 92% and Cr –> Pl:43%). The students' effort into the planning activity can be reflected in both the interactions and the amount of time they spent on planning (16%). However, because students' regulation levels and strategies varied, we could observe in the graph that in some instances (28%), the students jumped directly to the creation components.

Students' behaviours in condition C show that 80% of the content creation started with the creation activity. The strong interactions between the creation and monitoring (Cr –> Em: 60% and Em –> Cr: 90%) indicate the students used the monitoring support during the actual item writing process. However, there was a large probability of ending the process after writing the resource (40%). This might indicate the students did not use the monitoring support at all, used the monitoring support while writing or had already internalised the principles from their previous creation. The students spent only 4% of the total time on the scaffolding, which is reasonable, as the students did not need to take much time to read the hints. This observation is similar to previous research in which learners were found not to spend much time on monitoring, yet they very frequently engaged in it [11]. As expected, students in condition D started the task with the creation activity, then completed the self-assessment (88%) and submitted the created resource (61%). However, strong interactions between the self-assessment and creation were observed (Cr –> Sa: 88% and Sa –> Cr: 39%), indicating the students could review and revise their resources after self-assessment. These interactions made up 11% of the total creation time. Finally, students in condition E followed a combination of the processes observed in the other groups. However, the complexity of the interactions was increased as the intervention components could interact with one another. This complexity can be seen in the transitions from monitoring to planning (Em –> Pl:

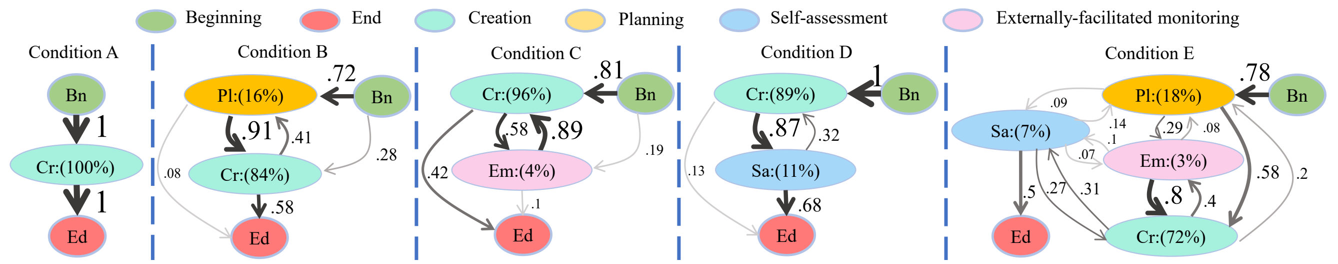

Figure 2: Overview of the processes and the time taken for creating content in each condition.

29%) and from self-assessment to planning (Sa –> Pl: 13%). The transition from monitoring to planning might indicate that sometimes the students reconsidered their plan based on the hints.

We examined the difference in the control group's time (min) against the treatment groups. Interestingly, the addition of the monitoring strategy (condition C) reduced the median time for creating a resource, but the change was not significant (Mdn = 11.2, U = 74907, p = .38). The addition of the self-assessing (condition D) scaffold increased the required time for creating a resource, but the additional time was not statistically significant (Mdn = 13.3, U = 73429, p = .18). On the other hand, the addition of the planning (condition B) and all of the SRL strategies (condition E) significantly increased the required time for creating a resource (Mdn = 15.7, U = 53433, r = .11, $p < .05$) and (Mdn = 19.4, U = 82481, r = 23, $p < .01$) respectively.

In summary, the results show that students followed different processes to engage with the interventions. However, in general, the interventions added to the complexity of the task and required students to dedicate additional time.

RQ2: Impact on Student Perceptions. Table 2 provides descriptive statistics on students' perceptions of the effectiveness of each condition. Interestingly, students found the control interface the most effective (4.62 ± .61). Conditions C and D followed with 4.45 ± .74 and 4.43 ± .80, respectively. The two conditions that included the planner were perceived as the least effective but had the highest deviation in response, indicating a diversity of students' opinions on their effectiveness (condition B 4.16 ± .99 and condition E 4.25 ± 0.98). We only examined the difference between the medians of the control group (condition A) and condition B since the other treatment groups had a similar median effectiveness rating to those in the control group (Mdn = 5). A Mann-Whitney U test showed no statistically significant difference between the two conditions.

The results of the qualitative analysis are summarised in Table 2. The control group responses had a split between the technology and the generic code and mostly provided positive comments, as shown by the average sentiment scores. In all treatment cases, more than 50% of students' responses were about the technology. The addition of the planner in conditions B and E seems to have introduced technological issues, as shown by their sentiment average scores under technology. The comments point to two main issues. First, the planner had made the creation page more complex, which made the interface laggy (e.g., "Quit laggy when writing in questions and responses."). Second, the addition of the planner made

Table 2: Students' perceptions of effectiveness for each condition (quantitative and qualitative analysis)

			Control	Treatment			
			A	B	C	D	E
Rating	Level	Very high	23 %	14.9 %	18 %	24.2 %	19.9 %
		High	17.1 %	21 %	18.5 %	21.5 %	22 %
		Medium	10.7 %	23.8 %	15.5 %	25 %	25 %
		Low	0%	40 %	13.3 %	13.3 %	33.3 %
		Very low	0%	37.5 %	0%	12.5 %	50%
	μ		4.62	4.16	4.45	4.43	4.25
	σ		.61	.99	.74	.8	.98
	Mdn		5	4	5	5	5
	U		—	41107	60463	58248	55515
	p		—	.11	.3	.16	.02
	N		24	35	13	36	29
Theme	Technology	%	42	74	85	56	59
		S($\mu \pm \sigma$)	.43±.45	.04±.52	.45±.29	.46±.49	.03±.36
	Pedagogy	%	12	14	0	17	24
		S($\mu \pm \sigma$)	.65±.03	.32±.21	—	.26±.26	.3±.31
	Generic	%	46	11	15	28	17
		S($\mu \pm \sigma$)	.71±.1	.63±.25	.5±0	.69±.21	.64±.38

it significantly longer to create a resource (see the results of RQ1), increasing the need for students to have the ability to save a draft of their resource, which was not supported by the system (e.g., "Does not save drafts. Had to recreate the whole question after accidentally clicking on the side bar."). The addition of monitoring in condition C and self-assessment in condition D seems to have been more positively viewed (e.g., "Very easy to follow lay out, lots of text box options.". However, student responses show that these strategies also introduced some technological difficulties (e.g., "RiPPLE is extremely easy to use, however, it may be a little glitchy at times"). Moreover, there were examples of students questioning the effectiveness of the interventions. For example, "Make it with fewer steps" and "self-reflection is fundamentally useless at the stage of drafting the question - should only ask for feedback after other people have assessed it." are examples of students in conditions E and D doubting the benefits of the pedagogical interventions.

In summary, students in the control group seem to have found the creation interface more effective than students in the experimental groups. Comments from the students suggest that the addition of the interventions, particularly the planner, introduced technological issues that were not anticipated in the design of the experiment.

RQ3: Impact on Performance. The Figure 3 all students category shows that conditions B (Mdn = 3.9), C (Mdn = 3.9) and D (Mdn = 3.85) had slightly higher median values for the quality rating of their learning products than the control group (Mdn = 3.8), while condition E had a similar median value (Mdn = 3.8). In all cases, the difference was not statistically significant ($p > .05$).

All Students	Stat	Control	Treatment			
		A	B	C	D	E
	Mdn	3.80	3.90	3.90	3.85	3.80
	q1	3.40	3.30	3.40	3.50	3.30
	q3	4.10	4.10	4.10	4.10	4.10
	U	-	69484	84185	86886	69803
	p	-	.94	.61	.59	.12
	r	-	.00	.02	.02	-.06

Engaged	Stat	Control	Treatment			
		A	B	C	D	E
	Mdn	3.80	3.90	3.90	3.90	3.80
	q1	3.40	3.50	3.70	3.50	3.30
	q3	4.10	4.20	4.20	4.10	4.10
	U	-	37948	25719	84103	14560
	p	-	.15	.007	.43	.88
	r	-	.06	.11	.03	-.01

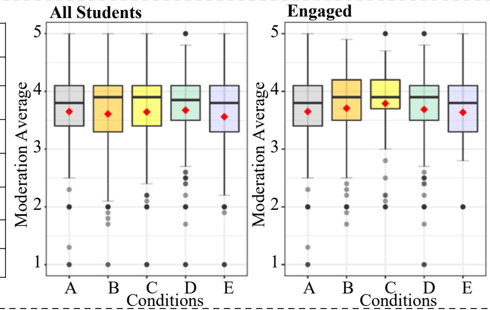

Figure 3: Overview of the quality rating of resources for all students and engaged students across control and treatment conditions.

When considering the engaged student category, Figure 3 shows that conditions B, C and D still had higher median values but also higher first quartiles for the three conditions (B: $Q1$ = 3.5, C: $Q1$ = 3.7, D: $Q1$ = 3.5) and higher third quartiles for conditions B and C (B: $Q3$ = 4.2, C: $Q3$ = 4.2). Condition E quality rating was about the same among all students and the engaged students, showing no difference to the control group. The Mann-Whitney U test and Pearson's r showed that the difference between the intervention (condition C) and the control (condition A) for the engaged students was statistically significant with a small effect (U = 25719, r = .11, p < .01). Nevertheless, the values were not significantly different from the control group for the other conditions.

4 DISCUSSION & CONCLUSION

As learning analytics (LA) is maturing, there is a growing interest in creating analytics-based environments that enhance learning and using learning analytics to evaluate such environments. This paper contributes to both of these aims by investigating the challenges and opportunities of implementing self-regulation in a large-scale learnersourcing environment. We observed the impact of technological choices on students' perceived benefit of and experience with the self-regulation interventions we implemented in the learning platform.

Our results show that adding metacognitive scaffolds increased task complexity and completion time. This result is not surprising and aligns with previous work [7] as the interventions require students to dedicate extra time to completing additional tasks. While effort put towards regulation can be a predictor of learners' grades [7], in our case, the additional effort did not contribute to significant improvements in performance, which has also been observed in previous work [4, 6, 25]. Many factors can help explain why the SRL strategies might have a small or no impact on content creation. For the engaged students, the small impact we observed might occur because those learners are stronger regulators than most of their peers [8]. For all students, low prior knowledge may have affected the use of the regulation prompts [6] and reduced their impact. Additionally, the technical difficulties that the students reported during learning might have mitigated the effects of the interventions; this was evident among conditions B and E, emphasising the importance of having the proper technology before pedagogy. Given that many students strategically focus mostly on tasks that contribute to their final grade [36], a potent challenge for SRL direct instruction interventions is to identify ways of providing support without requiring much additional time and effort.

Our results also show that the metacognitive scaffolds were not necessarily assessed positively by learners. According to Winne [35], for learners to use a learning tool that introduces metacognitive scaffolds, four conditions need to be met: - a) learners are aware that the scaffold is useful to them, b) learners know that the scaffold can be useful for the task at hand, c) learners have enough skills to use the scaffold and d) learners are motivated to use it. Based on the responses of the students and their low appreciation of the scaffolds (RQ2) – i.e., perceived usefulness of a tool is probably the best measure of metacognitive knowledge (related to conditions A-C) [9], it seems that students participating in our study might not have gained sufficient metacognitive awareness, knowledge and skills to use and engage productively with the scaffolds. This is where LA research, as discussed below, can contribute to addressing the issue.

Implications for the role of LA in supporting SRL. Along with the existing implementation strategies in the literature (e.g., [28]), our study highlights the need for combining newly introduced scaffolds with rationale for their value along with training to facilitate student acceptance and adoption. Here, we illuminate the role of LA in assessing the use of the scaffolds and providing personalised feedback on students' use of the tool and how they can use them more effectively in large-scale systems during and after the training. Importantly, this should be done over the course of several tasks - e.g., the learnersourcing tasks we used in the current study could be divided into several smaller tasks that are tailored with LA analytics and feedback on learners' use of the scaffolding tools. We also highlight the critical role of empirical research, and in particular control experiments, in measuring the effectiveness of theory-based technological interventions. While due to their theoretical framing, there is strong reason to believe that they would be effective, there are often many technological side effects, as was the case in our study, that limit their effect. Finally, we stress the importance of providing avenues for stakeholders to provide feedback about their experience as they interact with LA systems. In our case, it would have been challenging to detect the technological issues with the interventions without direct feedback from the students.

Limitations and future work. We see four main limitations to the current study. First, the duration of the experiment was relatively short, which may not have allowed the benefits of the interventions to be fully achieved. Accordingly, future work could address this limitation by repeating the experiment for a longer duration. Second, the actual quality of the resources was approximated via a peer review process rather than by experts, which may be prone to error. Accordingly, future work could address this limitation by having the quality of resources quantified by experts. Third, our results do not investigate the impact of the interventions on student learning (learning gains). An interesting future direction would be to conduct pre- and post-tests on the topics a student creates content in to explore the causal effect of the interventions on learning. Fourth, the experiment considered data from only two courses, which may restrict the generalisability of the presented findings. Future work aims to replicate the study with participants from over 10 of the courses that have currently adopted RiPPLE.

REFERENCES

[1] Solmaz Abdi, Hassan Khosravi, and Shazia Sadiq. 2020. Modelling Learners in Crowdsourcing Educational Systems. In *International Conference on Artificial Intelligence in Education*. Springer, 3–9.

[2] Shaaron Ainsworth. 2006. DeFT: A conceptual framework for considering learning with multiple representations. *Learn Instr* 16, 3 (2006), 183–198.

[3] Roger Azevedo and Vincent Aleven. 2013. *International handbook of metacognition and learning technologies* (2013 ed.). Springer international handbooks of education, Vol. 26. Springer, New York, NY.

[4] Roger Azevedo and Allyson F. Hadwin. 2005. 7g and Metacognition – Implications for the Design of Computer-based Scaffolds. *Instr. Sci.* 33, 5-6 (Nov. 2005), 367–379. Number: 5-6.

[5] Roger Azevedo, Daniel C Moos, Jeffrey A Greene, Fielding I Winters, and Jennifer G Cromley. 2008. Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development* 56, 1 (2008), 45–72.

[6] Maria Bannert and Christoph Mengelkamp. 2013. Scaffolding hypermedia learning through metacognitive prompts. In *International handbook of metacognition and learning technologies*. Springer, New York, NY, 171–186.

[7] Jaclyn Broadbent, Ernesto Panadero, Jason M Lodge, and Paula de Barba. 2020. Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In *Handbook of research in educational communications and technology*. Springer, Cham, 37–52.

[8] Jaclyn Broadbent and Walter L Poon. 2015. Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education* 27 (Oct 2015), 1–13.

[9] Geraldine Clarebout, Jan Elen, Norma A Juarez Collazo, Griet Lust, and Lai Jiang. 2013. Metacognition and the use of tools. In *International handbook of metacognition and learning technologies*. Springer, New York, NY, 187–195.

[10] Ali Darvishi, Hassan Khosravi, and Shazia Sadiq. 2021. Employing Peer Review to Evaluate the Quality of Student Generated Content at Scale: A Trust Propagation Approach. In *Proceedings of the Eighth ACM Conference on Learning@ Scale*. 139–150.

[11] Victor M Deekens, Jeffrey A Greene, and Nikki G Lobczowski. 2018. Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology* 88, 1 (Mar 2018), 63–79.

[12] Steven M. Downing. 2006. Twelve Steps for Effective Test Development. In *Handbook of test development*. Lawrence Erlbaum Associates Inc, Mahwah, N.J, 3–26.

[13] Dragan Gašević, Olusola Adesope, Srećko Joksimović, and Vitomir Kovanović. 2015. Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *Internet High Educ* 24 (Jan 2015), 53–65.

[14] Dragan Gašević, Shane Dawson, Tim Rogers, and Danijela Gasevic. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education* 28 (Jan 2016), 68–84.

[15] Dragan Gašević, Negin Mirriahi, Shane Dawson, and Srećko Joksimović. 2017. Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior* 67 (Feb 2017), 207–220.

[16] Roberto Gatta, Jacopo Lenkowicz, Mauro Vallati, Eric Rojas, Andrea Damiani, Lucia Sacchi, Berardino De Bari, Arianna Dagliati, Carlos Fernandez-Llatas, Matteo Montesi, et al. 2017. pMineR: an innovative R library for performing process mining in medicine. In *AIME*. Springer, 351–355.

[17] Jeffrey Alan Greene and Roger Azevedo. 2009. A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology* 34, 1 (Jan 2009), 18–29.

[18] Jasmina Jovanovska. 2018. Designing Effective Multiple-Choice Questions for Assessing Learning Outcomes. *Infotheca* 18, 1 (Sep 2018), 25–42.

[19] Hassan Khosravi, Gianluca Demartini, Shazia Sadiq, and Dragan Gasevic. 2021. Charting the Design and Analytics Agenda of Learnersourcing Systems. In *11th International Learning Analytics and Knowledge Conference*. ACM, 32–42.

[20] Hassan Khosravi, Kirsty Kitto, and Williams Joseph. 2019. RiPPLE: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities. *Journal of Learning Analytics* 6, 3 (2019), 91–105.

[21] Hassan Khosravi, Shazia Sadiq, and Dragan Gasevic. 2020. Development and adoption of an adaptive learning system: Reflections and lessons learned. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 58–64.

[22] Juho Kim. 2015. *Learnersourcing: improving learning with collective learner activity*. Thesis. Massachusetts Institute of Technology.

[23] Xiaodong Lin, Cindy Hmelo, Charles K. Kinzer, and Teresa J. Secules. 1999. Designing technology to support reflection. *Educational Technology Research and Development* 47, 3 (Sept. 1999), 43–62.

[24] Aaron McDonald, Heath McGowan, Mollie Dollinger, Ryan Naylor, and Hassan Khosravi. 2021. Repositioning students as co-creators of curriculum for online learning resources. *Australasian Journal of Educational Technology* (2021), 102–118.

[25] Inge Molenaar, Carla A. M. van Boxtel, and Peter J. C. Sleegers. 2011. Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science* 39, 6 (Oct 2011), 785–803.

[26] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (Apr 2017), 422.

[27] Ernesto Panadero, Anders Jonsson, and Juan Botella. 2017. Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review* 22 (Nov 2017), 74–98.

[28] Ernesto Panadero, Julia Klug, and Sanna Järvelä. 2016. Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research* 60, 6 (Nov 2016), 723–735.

[29] Katherine A Rawson and John Dunlosky. 2013. Retrieval-monitoring-feedback (RMF) technique for producing efficient and durable student learning. In *International handbook of metacognition and learning technologies*. Springer, New York, NY, 67–78.

[30] Tyler W Rinker. 2017. Sentimentr: Calculate Text Polarity Sentiment. Version 1.0. 0.

[31] Ido Roll and Philip H. Winne. 2015. Understanding, evaluating, and supporting self-regulated learning using learning analytics. *JLA* 2, 1 (July 2015), 7–12.

[32] John Saint, Yizhou Fan, Shaveen Singh, Dragan Gasevic, and Abelardo Pardo. 2021. Using Process Mining to Analyse Self-Regulated Learning: A Systematic Analysis of Four Algorithms. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21)*. ACM, 333–343.

[33] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *Am J Eval* 27, 2 (Jun 2006), 237–246.

[34] Joep van der Graaf, Lyn Lim, Yizhou Fan, Jonathan Kilgour, Johanna Moore, Maria Bannert, Dragan Gasevic, and Inge Molenaar. 2021. Do Instrumentation Tools Capture Self-Regulated Learning?. In *11th International Learning Analytics and Knowledge Conference (LAK21)*. ACM, 438–448.

[35] Philip H Winne. 2006. How software technologies can improve research on learning and bolster school reform. *Educ. Psychol.* 41, 1 (Mar 2006), 5–17.

[36] Philip H. Winne. 2013. Learning Strategies, Study Skills, and Self-Regulated Learning in Postsecondary Education. In *Higher Education: Handbook of Theory and Research*, Michael B. Paulsen (Ed.). Higher Education: Handbook of Theory and Research, Vol. 28. Springer, Dordrecht, 377–403.

[37] Barry J Zimmerman and Adam R Moylan. 2009. Where metacognition and motivation intersect. *Handbook of metacognition. Routledge* (2009), 299–316.