

The Effects of Reflective Pauses on Performance in Simulation Training

Joy Y. Lee, PhD;

Jeroen Donkers, PhD;

Halszka Jarodzka, PhD;

Géraldine Sellenraad, MD;

Tjitske J.E. Faber, PhD;

Jeroen J.G. van Merriënboer, PhD

Introduction: The reflective pause, taking a pause during performance to reflect, is an important practice in simulation-based learning. However, for novice learners, it is a highly complex self-regulatory skill that cannot stand alone without guidance. Using educational theories, we propose how to design cognitive and metacognitive aids to guide learners with the reflective pause and investigate its effects on performance in a simulation training environment.

Methods: These effects are examined in four aspects of performance: cognitive load, primary performance, secondary performance, and encapsulation. Medical students ($N = 72$) performed tasks in simulation training for emergency medicine, under 2 conditions: reflection condition ($n = 36$) where reflection was prompted and guided, and control condition ($n = 36$) without such reflection.

Results: The effects of reflective pauses emerged for 2 aspects of performance: cognitive load decreased and secondary performance improved. However, primary performance and encapsulation did not show significant difference.

Conclusions: The results demonstrate that reflective pauses with cognitive and metacognitive aids implemented can enhance some aspects of performance. We suggest that to secure these effects, feedback during reflection and an adaptation period should be provided.

(*Sim Healthcare* 19:82–89, 2024)

Key Words: Self-reflection, pausing, simulation-based training, serious game, cognitive load, eye tracking.

In critical fields, such as surgery and emergency medicine, the tasks are extremely dynamic and high stakes. In these tasks, practitioners should constantly adapt their performance by making a series of decisions on demand.^{1,2} To make these decisions successful, *the reflective pause*,³ stopping the current course of action for a brief time to reorganize their performance, becomes an essential strategy. Studies have reported the reflective pause may improve performance and learning.^{4–9}

However, taking pauses to reflect on one's own performance is a highly complex self-regulatory skill, which requires guidance especially for novices.^{9,10} It involves metacognitive processes to monitor and regulate own performance,¹¹ including clinical judgment and time management.¹² It is well known that students are not competent in these metacognitive self-regulatory skills.^{13–15} In an empirical study, Lee et al¹⁶ demonstrated that students struggle with when and how to

take pauses and that allowing for learner-initiated pauses without guidance results in limited contribution to their performance. Researchers in education have stressed that, to acquire these metacognitive skills, learners should be provided with systematic support and guidance.^{17,18}

To support novice learners with reflective pauses, we propose how to design cognitive and metacognitive aids (CMAs).¹⁸ According to theories from complex learning, CMAs for reflective pauses can involve the following 3 elements: prompts, cues, and leading questions.^{18–20}

Prompts

Prompts support when to pause within scenarios. To systematically deploy pauses with right timing, scenarios can be segmented into smaller parts, while pauses are inserted between these segments to facilitate the comprehension of the content structure.^{21,22} A good example of this segmentation is parttask sequencing methods,¹⁸ which creates clusters of part tasks and arranges them in sequence.

Cues

Once a pause is prompted at the right time, what to reflect on is the next challenge. A well-known pitfall of reflection is its dependency on episodic memories, which often leads to forgetting or fabrication.^{23,24} To compensate for this drawback, records of learners' performance (ie, cues) can be provided to support the memories about their performance.²⁵

Leading Questions

Learners tend to concentrate on surface-level features from the cues and lose learning opportunities to restructure their knowledge.^{26,27} To discover meaningful information from the cues, learners should be guided in how to reflect. Leading questions²⁸ are a good aid for this, which helps learners start

From the Faculty of Governance and Global Affairs (J.Y.L.), Leiden University, The Hague; School of Health Professions Education (J.D., J.J.G.v.M.), Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht; Faculty of Education Sciences (H.J.), Open Universiteit, Heerlen; Faculty of Health, Medicine and Life Sciences (G.S.), Maastricht University, Maastricht; and Institute for Medical Education Research Rotterdam (T.J.E.F.), Erasmus University Medical Center, Rotterdam, the Netherlands.

Correspondence to: Joy Y. Lee, PhD, Faculty of Governance and Global Affairs, Leiden University, P.O. Box 9500, 2300 RA Leiden, the Netherlands (e-mail: j.y.lee@luc.leidenuniv.nl).

The authors declare no conflict of interest.

Supported by the Netherlands Organization for Scientific Research (NWO, grant numbers 055.16.117).

Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the Society for Simulation in Healthcare. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. DOI: 10.1097/SIH.0000000000000729

TABLE 1. Demography of Each Group

	No RP	RP
Age, mean (SD), yr	22.5 (2.4)	22.0 (2.4)
No. female	27	26
No. extra experienced*	16	16
No. senior (4th to 6th year)	18	16
Total	36	36

*Extra experience includes internship in emergency department or training in simulation center.

from what they already know and develop their knowledge by self-explaining the relationships between pieces of information.^{20,29}

In this study, we posit that the reflective pause is a complex self-regulatory skill that can hardly be effective without CMAs. Accordingly, we examine the effects of the reflective pause with CMAs on performance, by using a simulation training for emergency medicine. To assess performance, we consider multiple aspects of performance, as combining different aspects of performance in a global manner allows for a well-designed assessment.^{30–32} We expect that the following 4 aspects of performance can be enhanced by the reflective pause: cognitive load, primary performance (diagnosis and intervention), secondary performance (vigilance), and encapsulation. Hypotheses are formed per aspect as described hereinafter.

Cognitive load has been widely used as an indicator of performance in the medical fields.^{33–35} The basic idea is that if working memory is cognitively overloaded, performance deteriorates. In a condition where the reflective pause is applied, overall cognitive load would decrease as participants optimize their working memory through reflection (Hypothesis 1, H1). Because this optimization creates more space for working memory, primary performance (diagnosis and intervention in a resuscitation task) is enhanced (H2), as well as secondary performance (vigilance on situational changes during the resuscitation) (H3). During the reflection, knowledge structure is activated to compose higher-order concepts.³⁶ This process of knowledge encapsulation allows for better clinical reasoning,^{37,38} resulting in better-summarized handover report (H4).

METHODS

Participants and Design

After approval from our institutional research ethics board (FHML-REC/2020/007/Amendment1), 72 medical students (53 females, mean age = 22.3, SD = 2.4) from Maastricht University, the Netherlands, voluntarily participated. The recruitment was from March 1, 2020, to March 30, 2021, being delayed because of COVID-19. The sample size is determined as it is widely used in eye-tracking research and validated by our previous study with similar setups.¹⁶ The participants' academic years varied from second to sixth. They all had taken the basic course of emergency medicine in their first year, while their experience in the domain varied according to their individual curriculum. Using a between-subjects design, a researcher (J.Y.L.) assigned participants into 2 groups by allocating randomly generated participation number: the experimental group where participants were prompted to reflect on their performance during pauses (RP condition, $n = 36$), and the control group where they were asked to perform an irrelevant

task instead during the same pauses (no-RP condition). Table 1 gives an overview of demography of each group. No blinding was involved. For all participants, we provided 3 pauses between parttasks grouped by skill clusters. Their performance in each cluster was repeatedly measured for each participant via the measures described in Table 2.

Materials and Apparatus

A Medical Simulation Game

We used a computer-based simulation game for emergency medicine, AbcdeSIM.³⁹ The 5 letters ABCDE stand for the 5 phases that should be followed for acute care: airway (A), breathing (B), circulation (C), disabilities (D), and exposure (E). This game simulates real-life situations in the emergency department, where a high-fidelity human physiology is programed to react to the learners' interaction. The gastrointestinal bleeding (GIB) scenario was used because its level of complexity is high enough to provoke salient cognitive load.³² This scenario presents a 32-year-old male patient with hypovolemic shock. The primary task is to stabilize this patient by diagnosing and intervening. The game score was calculated by summation of correct in-game actions of diagnosis and intervention, yielding the measure to test H2 (primary performance).

Cognitive and Metacognitive Aids

The prompts were implemented between parttask clusters. We translated the five phases of ABCDE into five skill clusters which were then composed into 3 task clusters: AB – ABC – ABCDE. These task clusters were designed using forward chaining,¹⁸ a useful part-task sequencing method for teaching medical interventions. A current physician in emergency medicine developed this clustering and time-on-task required for each cluster.

The cues included a gameplay recording where eye movements are superimposed on the gameplay screen, and a heatmap that shows eye fixation allocation (Fig. 1). The former is expected to show the learner's performance and attention allocation in a timeline, while the latter shows the overall attention allocation in summary. The game play and eye movements were recorded and displayed by SMI BeGaze software (version 3.6, www.smivision.com).

The leading questions asked during the pauses were five: “Did I miss something important to check?” “What did I check unnecessarily?” “Did I miss any important intervention?” “What redundant interventions have I applied?” and “How to improve my performance?” The participants were asked to verbalize their reflection (ie, think aloud) to ensure that they are involved in the activity.

TABLE 2. Hypotheses, Constructs, and Measures

Hypothesis	Construct	Measure
H1	Cognitive load	Pupil increase (mm) Paas scale (1–9)
H2	Primary performance (Diagnosis and intervention)	Game score
H3	Secondary performance (Vigilance on patient status)	Transition rate across the VSM (/min)
H4	Encapsulation	Handover report



FIGURE 1. Heatmap as a type of cognitive and metacognitive aid.

Eye Tracking

Eye-tracking data was collected by an SMI RED remote eye-tracker with a sampling rate of 250 Hz and SMI iView X software (version 2.7.13). We used a dual-computer setup where a laptop computer with the iView X software was connected to a personal computer for game presentation. We used SMI Experiment Center 3.5 software (version 3.2.11) to arrange calibration and presentation of instruction and stimuli. A forehead-and-chin rest was used to prevent head movements and make the geometry between the eye and the eye tracker stable.⁴⁰ The experiment room maintained a constant luminance, without windows or noise, composing a controlled environment for eye-tracking.

In the GIB scenario, the patient's vital signs change dynamically and extra attention to the vital signs monitor (VSM) is required. We define this vigilance on patient status as the secondary performance as it supports the primary performance. To detect this vigilance, eye-tracking is used where eye-fixation *locations produce transition rate*. Transition rate refers to the number of gaze shifts per second from one area of interest (AOI) to another, which can represent cognitive processes to collect information across different AOIs.⁴⁰ In medical contexts, it has been used to quantify vigilance or expertise

levels, showing a higher transition rate for experts.^{41–43} We defined the upper-middle area of the screen where the VSM is located as an AOI (Fig. 2). The transition rate between this VSM AOI and non-VSM areas was expected to reflect the vigilance on patients' status, allowing us to test H3 (secondary performance).

Cognitive Load Measures

To test H1 (cognitive load), we used 2 measures for a more comprehensive interpretation of cognitive load⁴⁴: pupillometry, an objective indicator of cognitive load in medical simulation settings,^{35,45} and Paas scale,⁴⁶ a subjective rating scale for cognitive load. Paas scale has 9-point ratings with the value 1 representing the lowest cognitive load and the value 9 representing the highest.

Handover Report

A handover report is a brief document where the patient case is communicated in a concise way. It requires knowledge encapsulation and clinical reasoning for given scenarios,⁴⁷ thus, we assume that it can provide a reasonable measure to test H4 (encapsulation). Two current physicians designed a checklist that consists of 17 items extracted from SBAR,⁴⁸ the internationally recognized protocol for clinical handover. These items included patient information (eg, age, medication, past medical history), diagnosis (eg, oxygen saturation, melena,

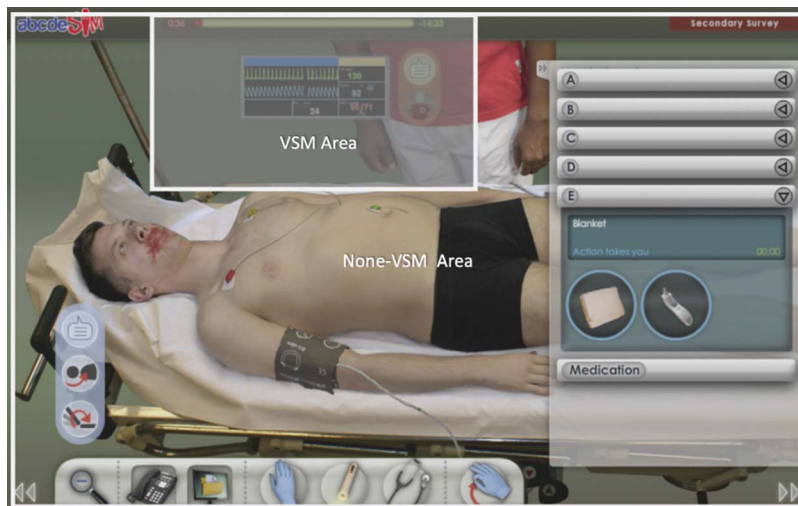


FIGURE 2. Area of interest definition: VSM area versus non-VSM area.



FIGURE 3. Procedure of a session.

tachypnea, tachycardia, laboratory results), and intervention (eg, oxygen administration, intravenous infusion, gastroscopy). They are selected as the most relevant parameters for GIB, which represent an efficient summary of the scenario. The physicians, then, assessed the accuracy and conciseness of participants' reports based on this list. For each item that is incorrectly summarized, one point was deducted. The physicians' scores were averaged, and the interrater reliability was examined ($r(69) = 0.98, P < 0.001$).

Procedure

Figure 3 shows the entire procedure of a session. The participants were individually invited to the eye-tracking laboratory. They signed an informed consent form and filled out a demographic questionnaire. In the pretraining, they watched a tutorial and played an easy scenario to familiarize themselves with the game functions. Participants in RP were additionally instructed in how to use the cognitive and metacognitive aids (CMAs) for their reflective pauses. All participants, then, were positioned for the eye-tracking setup and followed a 9-point calibration procedure. A scrambled image of the game screen was presented for 5 seconds to establish a baseline for pupillometry. Next, the participants played the first task cluster (2 minutes) with a timer visible on the screen. When the time was up, the first pause was prompted. During this pause, participants in RP reflected on their performance using the cues and the leading questions. During the same pause, the participants in no RP watched an advertisement video of an unrelated medical simulation game and assessed this game in writing. This task is expected to impose a task effort similar to that of the reflection task in RP. It is not related to self-reflection but still a cognitive evaluative process within medical domains, which should not be too distractive. For the second (4 minutes) and the third (6 minutes) task cluster, the same procedure from the calibration until the activities during the pauses was repeated. After each cluster, the participants rated their perceived cognitive load using the Paas scale. After the third pause, the participants filled in the handover report. The entire session took approximately 40 minutes.

Data Analysis

Eye-Tracking Data

The eye-tracking data included sampled measurements of pupil diameter and eye fixation location. These data were

merged with game logs by synchronizing the game system with the eye-tracking software and then imported into R version 3.5.1.⁴⁹ Pupil diameter data were processed following the method from Lee et al¹⁶ that refines eye-tracking data to eliminate confounding factors in pupil dilations. As a result, means of absolute pupil increase against the baseline were obtained for each cluster.

Statistical Analysis

To investigate the effects of reflective pauses across all clusters, we first fitted linear mixed-effects models, using the lme4 package⁵⁰ in R. "Condition" and "Cluster" were entered as fixed effects with interaction term, while "participant" was treated as a random factor (random intercept). To see whether the effects exist in specific clusters, we built separate models with dummy variables. For the effects of the first pause on the second cluster, we defined dummy variables for Cluster corrected for the difference at the first cluster. For the effects of the second pause on the third cluster, another dummy encoding for Cluster was used to correct for the difference at the second cluster.

For all models, residual plots were visually inspected and did not reveal any obvious deviations from homoscedasticity or normality. Likelihood ratio tests of the full model with the effects in question against the model without these effects obtained *P* values. R was used for all statistical analysis. We considered $P < 0.05$ to be statistically significant, except for the effects in specific clusters where $P < 0.025$ was applied to correct for the multiple comparisons.

RESULTS

Six clusters in the eye-tracking data and one cluster in the game logs were missing because of technical issues in the connection between the game system and the iViewX software. For the data quality of transition rate, eye-tracking data with tracking ratio below 50% and accuracy deviation larger than 2 degrees were excluded. After the exclusion, the average tracking ratio was 96.9%, and the average accuracy deviation was 0.9 degrees. According to the recording of the verbalization, all participants in RP were engaged to reflection during pauses.

Table 3 shows the descriptives of all measures of performance, with outcomes sorted by clusters. In general, pupil dilation increased over the duration of the task. Exceptionally, it decreased in the third cluster in RP, suggesting a reduced

TABLE 3. Descriptive of Performance Over Task Clusters

Construct	Measure	No RP						RP					
		Cluster 1		Cluster 2		Cluster 3		Cluster 1		Cluster 2		Cluster 3	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Cognitive load	Pupil increase (mm)	0.80	0.29	1.01	0.25	1.02	0.24	0.97	0.24	1.08	0.26	0.84	0.34
	Paas scale* (1–9)	6	1.75	6	1.73	7	1.77	6	1.15	7	1.13	7	1.19
Primary performance	Game score	308	41.8	387	67	473	77.2	289	36.1	377	54.8	475	60
Secondary performance	Transition rate (/min)	9.55	6.4	12.5	5.14	11.7	4.26	10.5	5.83	11.8	3.9	14.6	6.47
Encapsulation	Handover report	5.4 (2.81)						5.64 (1.59)					

*Median was used instead of mean as a central tendency.

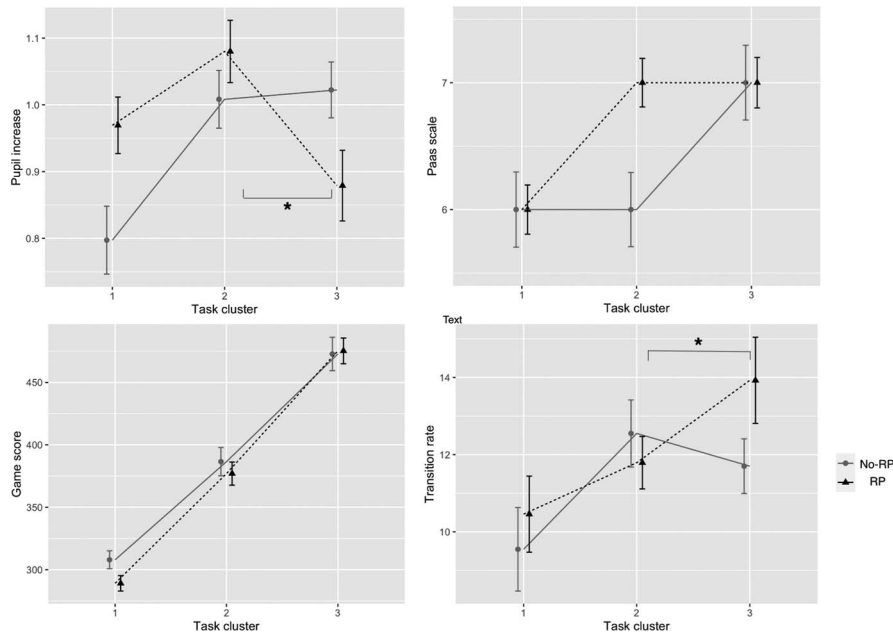


FIGURE 4. Performance change over task clusters. Note: Significant effects are marked with asterisks, which existed in 2 measures: pupil increase and transition rate. These effects emerged only in the third cluster with the reference to the second cluster.

cognitive load. The Paas scale resulted in high scores overall (ie, between 6 and 7) regarding the 9-point ratings. Game score increased throughout all clusters. Transition rate also increased over the clusters, except that it decreased at the third cluster in no RP, indicating lower vigilance on patient status. Handover report score was somewhat higher in RP. Figure 4 visualizes these changes in the measures.

To examine statistical significance in the effects of reflective pauses, we used linear mixed-effects models. When comparing RP against no RP in general, the models did not show any significant difference between the 2 conditions in any of the measures. However, when looking at the separate clusters, the third cluster showed a significant difference in 2 measures: pupil increase and transition rate. Table 4 presents the effects of condition at the second and the third clusters on the different measures.

To test H1 (cognitive load), we used pupil increase and the Paas scale. Pupil increase showed a significant difference between the conditions at the third cluster, while the Paas scale did not show any significant difference. For H2 (primary performance), the game score did not show any significant effects. The measure for H3 (secondary performance) demonstrated a significant effect in the expected direction, again at the third

cluster. No significant effects were detected in the handover report measure for H4 (encapsulation), although the difference was in the expected direction. The significant effects are marked with asterisks in Figure 4.

In addition, to see whether any difference exists in the reference (ie, baseline), a *t* test was used for each measure in the first cluster. The results of this test are expected to show whether the pretraining influenced the baseline of each measure. There was a significant difference of the condition only in pupil increase: It was higher in RP ($M = 0.97$, $SD = 0.24$) than in no RP ($M = 0.80$, $SD = 0.29$) ($t(63) = -2.59$, $P = 0.012$, $d = 0.64$, 95% CI = -0.30 to -0.04).

DISCUSSION

This study investigated the effects of the reflective pause on different aspects of performance in simulation training. We start from the ground that, for novice learners, taking reflective pauses is a highly complex self-regulatory skill, which cannot stand alone without guidance. Accordingly, we developed and provided reflective pauses with the following 3 types of CMAs: prompts, cues, and leading questions. We prompted 3 times for pauses between the 3 partask clusters and handover reporting. Of the 4 aspects of performance that we

TABLE 4. Effects of Reflective Pauses on the Measures of Cognitive Load and Performance

Construct	Measure	Effect	β	SE	<i>t</i>	<i>P</i>
CL	Pupil increase	Condition (RP) in cluster 2	-0.077	0.066	-1.160	0.248
		Condition (RP) in cluster 3	-0.245	0.066	-3.717	0.000
	Paas scale	Condition (RP) in cluster 2	0.027	0.254	0.107	0.915
		Condition (RP) in cluster 3	0.011	0.254	0.045	0.964
Primary performance	Game score	Condition (RP) in cluster 2	11.760	15.930	0.738	0.462
		Condition (RP) in cluster 3	-11.759	15.926	-0.738	0.462
Secondary performance	Transition rate	Condition (RP) in cluster 2	-1.717	1.826	-0.941	0.349
		Condition (RP) in cluster 3	4.254	1.850	2.299	0.024
Encapsulation	Handover report	Condition (RP)	-0.154	0.613	-0.252	0.802

Significant effects ($P < 0.05$) are in boldface. Cluster 1 was the reference for the cluster 2 effect, and cluster 2 was the reference for the cluster 3 effect.

examined, 2 aspects (ie, cognitive load and secondary performance) improved through reflective pauses. These effects appeared only in the third cluster in the parttask sequence.

The first hypothesis (H1) assumed that the reflective pause decreases cognitive load in the intense scenario, as the participants optimize their working memory through reflection. This was supported by pupillometry, showing a smaller pupil increase in the RP condition, while the subjective ratings did not show any significant difference between the conditions. We consider H1 to be at least partly confirmed because physiological measures such as pupillometry tend to show higher sensitivity in measuring cognitive load than subjective ratings, in simulation training where its task environment is highly dynamic.^{16,45}

Interestingly, the effects of reflective pauses existed only in the third and last parttask cluster, signifying some latency in the effects. We postulate 2 explanations for this latency. First, an adaptation period is required to learn how to reflect using the CMAs. Metacognitive activities such as self-reflection are challenging for novices,^{16,51} and participants were required to learn how to use the CMAs and process the novel information from them. It is likely that the participants were having trouble in the earlier stages to adapt themselves to the new tools. From the additional analysis on the baseline of each measure, we discovered significantly higher pupil increase at the first cluster in RP. We deem that this result indicates the participants' challenge against the adaptation, foreshadowing the latency.

The second explanation for the latency stems from the dynamic nature of tasks in critical fields. Task environments in these fields are continuously changing, where novel information gradually emerges as time progresses. The problems in these tasks are rarely caused by a single factor but are an accumulation of minor and latent errors.⁵² Naturally, practitioners tend to first focus on task completion itself rather than reflecting on performance, while the signals of errors accumulate to a tipping point that composes an integral perception of a problem.^{53,54} It is probable that information necessary for reflection was not fully accumulated for the participants at the earlier stage, resulting in a lack of significant effects. This latency should be studied further in the future, for instance, by testing with a lengthened training duration where more time is provided to establish the effects on performance.

The second hypothesis (H2) expected the reflective pause to improve the primary performance (diagnosis and intervention). Contrary to our expectation, there was no significant difference between RP and no RP in any cluster. We suppose that the students could not arrive at the development of solutions to improve their future performance, because of the lack of feedback. Feedback is the information that indicates whether points of improvement exist.⁵⁵ The students might have been struggling with identifying the gap between their problem states and goal states, not knowing how to improve their performance. Bardach et al⁵⁶ reported that combining reflection and feedback is more effective than using only either reflection or feedback, which is in line with our explanation. Although we intentionally excluded the feedback process from our study to identify the exact effects of reflection, future research should compare this setup to that with feedback. In addition, more sensitive measure of domain-specific performance than game score should be explored further.^{30,31}

The third hypothesis (H3) assumed that the reflective pause enhances the secondary performance (vigilance). In the RP, the transition rate across the VSM was significantly higher in the third cluster, supporting H3's assumptions. We postulate that as the optimization of cognitive load creates additional space for SL, participants could better attend to situational changes, enhancing the vigilance on patient status. Transition rate has been known as an effective measure of vigilance or situational awareness in diverse medical domains.^{32,42,43} Our findings demonstrate that it can also measure behavioral change in visual attention in simulation training. Besides, we consider that presenting participants' eye movements as a cue for reflection fostered this change. Future research could further investigate how to use the visual cues as an educational technique to facilitate reflection.

The last hypothesis (H4) assumed the improvement in encapsulation thanks to the activation of knowledge structure and composition of higher-order concepts during reflection. H4 was not corroborated, probably because of similar reasons as for H2. Because of the lack of feedback, participants could not properly summarize the scenario nor identify points of improvement. Thus, despite the knowledge structure activation, higher-order conceptualization was not formed, resulting in no improvement in encapsulation. We suggest future studies to take a lengthened period to find the longitudinal effects on encapsulation, as the observed difference was in the expected direction.

This study opens new possibilities for research on in-action reflection in simulation-based learning. Based on the assumption that novice learners should be guided for an effective reflection, we have demonstrated an example of how to design CMAs. There is no one good CMA that fits all different task environments. To make reflective pauses effective, proper CMAs should be developed through task analysis and specification of learning goals. Using multiple CMAs and performance measures, we have provided evidence for the benefits of the reflective pause with CMAs implemented. Its positive effects on managing heightened cognitive load and increasing vigilance are clear advantages for professional skills in critical fields. Based on our findings, future research can be developed by investigating relevant topics: the effects of CMAs on learning, combination of reflection and feedback, and instructional design that integrates the reflective pause.

From a practical view, designers of computer-based simulation systems can immediately implement the reflective pause in their systems by applying prompts, cues, and leading questions. To support novice learners who lack metacognitive skills, the systems can provide these CMAs to guide them through reflective pauses. Even without any feedback, the use of reflective pauses can already decrease learners' cognitive load in extreme scenarios and encourage learners to be circumspect on situational changes. However, some precautions should be considered: familiarizing students with reflection techniques takes time and effort, and their expected positive effects probably only show up in the later stage of learning.

This study has several limitations. First, although we had built the theoretical background for the effects of CMAs, the effects of each CMA were not independently studied. Second, the quality of reflection (eg, analysis of the think aloud during reflection) and its correlation with performance and learning

was not investigated. Third, we used one scenario in emergency medicine, customizing CMAs for this particular scenario. To generalize our findings, development of other CMAs specialized to different scenarios should follow. Fourth, because our experiment was conducted in a controlled laboratory environment, more factors from reality might have been excluded. For instance, time pressure during pauses can play a significant role in real situations.

CONCLUSIONS

To our knowledge, this study is the first attempt to identify tangible effects of the reflective pause in simulation-based training in critical domains. Reflection during performance in extreme task environments can seem to be contradictory and uncomfortable for practitioners. When facing an emergency, they first jump into the task and even a brief moment of reflection can be seen as a “luxury.” Therefore, researchers and educators should even more stress the benefits of the reflective pause, which can enhance performance and safety level. Moreover, the reflective pause can create learning opportunities to bolster lifelong learning for professionals, if implemented within a well-designed simulation training with enriched cognitive and metacognitive support.

ACKNOWLEDGMENT

The authors thank Dr Jeroen Reijnders for the performance assessment and Dr Shahab Jolani for the advice on data analysis.

REFERENCES

1. Schmutz JB, Lei Z, Eppich WJ, Manser T. Reflection in the heat of the moment: the role of in-action team reflexivity in health care emergency teams. *J Organ Behav* 2018;39(6):749–765.
2. Ishak AW, Ballard DI. Time to re-group: a typology and nested phase model for action teams. *Small Group Research* 2012;43(1):3–29.
3. Clapper TC, Leighton K. Incorporating the reflective pause in simulation: a practical guide. *J Contin Educ Nurs* 2020;51(1):32–38.
4. McMullen M, Wilson R, Fleming M, et al. “Debriefing-on-Demand”: a pilot assessment of using a “pause button” in medical simulation. *Simul Healthc* 2016;11(3):157–163.
5. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc* 2007;2(2):115–125.
6. Altpeter T, Luckhardt K, Lewis JN, Harken AH, Polk HC Jr. Expanded surgical time out: a key to real-time data collection and quality improvement. *J Am Coll Surg* 2007;204(4):527–532.
7. Trowbridge RL. Twelve tips for teaching avoidance of diagnostic errors. *Med Teach* 2008;30(5):496–500.
8. Rall M, Glavin R, Flin R. The ‘10-seconds-for-10-minutes principle’. *Bull R Coll Anaesth*. 2008;51:2,614–2,616.
9. Lee JY, Szulewski A, Young JQ, Donkers J, Jarodzka H, Van Merriënboer JGG. The medical pause: importance, processes and training. *Med Educ* 2021;55(10):1152–1160.
10. St-Martin L, Patel P, Gallinger J, Moulton CA. Teaching the slowing-down moments of operative judgment. *Surg Clin North Am* 2012;92(1):125–135.
11. Nelson TO. Metamemory: a theoretical framework and new findings. *Psychol Learn Motiv*. 1990;26:125–173.
12. Moulton CA, Regehr G, Mylopoulos M, Macrae HM. Slowing down when you should: a new model of expert judgment. *Acad Med* 2007;82(suppl 10):S109–S116.
13. McCabe J. Metacognitive awareness of learning strategies in undergraduates. *Mem Cognit* 2011;39(3):462–476.

14. Foerst NM, Klug J, Jöstl G, Spiel C, Schober B. Knowledge vs. action: discrepancies in university students’ knowledge about and self-reported use of self-regulated learning strategies. *Front Psychol* 2017;8:1288.
15. Finn B, Tauber SK. When confidence is not a signal of knowing: how students’ experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educ Psychol Rev* 2015;27(4):567–586.
16. Lee JY, Donkers J, Jarodzka H, Sellenraad G, Van Merriënboer JGG. Different effects of pausing on cognitive load in a medical simulation game. *Comput Human Behav* 2020;110:106385.
17. Schmutz JB, Kolbe M, Eppich WJ. Twelve tips for integrating team reflexivity into your simulation-based team training. *Med Teach* 2018; 40(7):721–727.
18. Van Merriënboer JJ, Kirschner PA. *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design*. New York: Routledge; 2018.
19. Butler DL, Winne PH. Feedback and self-regulated learning: a theoretical synthesis. *Rev Educ Res*. 1995;65(3):245–281.
20. Chiu JL, Chi MT. Supporting self-explanation in the classroom. In: Benassi VA, Overson CE, Hakala CM, eds. *Applying Science of Learning in Education: Infusing Psychological Science Into the Curriculum*. 2014:91–103. Available at: <http://teachpsych.org/ebooks/asle2014/index.php>. Accessed May 23, 2022.
21. Spanjers IAE, Van Gog T, Wouters P, Van Merriënboer JGG. Explaining the segmentation effect in learning from animations: the role of pausing and temporal cueing. *Comput Educ* 2012;59(2):274–280.
22. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. *Educ Psychol* 2003;38(1):43–52.
23. Russo JE, Johnson EJ, Stephens DL. The validity of verbal protocols. *Mem Cognit* 1989;17(6):759–769.
24. Kuusela H, Paul P. A comparison of concurrent and retrospective verbal protocol analysis. *Am J Psychol* 2000;113(3):387–404.
25. Van Someren M, Barnard Y, Sandberg J. *The Think Aloud Method: A Practical Approach to Modeling Cognitive Processes*. (Knowledge-based systems) London: Academic Press; 1994.
26. Sweller J, Kirschner PA, Clark RE. Why minimally guided teaching techniques do not work: a reply to commentaries. *Educ Psychol* 2007;42(2):115–121.
27. Kirschner F, Paas F, Kirschner PA. Individual and group-based learning from complex cognitive tasks: effects on retention and transfer efficiency. *Comput Human Behav* 2009;25(2):306–314.
28. Collins A, Ferguson W. Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educ Psychol* 1993;28(1):25–42.
29. Renkl A. Worked-out examples: instructional explanations support learning by self-explanations. *Learn Instruct* 2002;12(5):529–556.
30. Cunnington JP, Neville AJ, Norman GR. The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ Theory Pract* 1996;1(3):227–233.
31. Dankbaar ME, Stegers-Jager KM, Baarveld F, et al. Assessing the assessment in emergency care training. *PLoS One* 2014;9(12):e114663.
32. Lee JY, Donkers J, Jarodzka H, van Merriënboer JJ. How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Comput Human Behav*. 2019;99:268–277.
33. Sweller J, Van Merriënboer JJ, Paas F. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev*. 2019;31(2):261–292.
34. Szulewski A, Howes D, Van Merriënboer JGG, Sweller J. From theory to practice: the application of cognitive load theory to the practice of medicine. *Acad Med* 2021;96(1):24–30.
35. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti MLA, van Merriënboer JGG. Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Adv Health Sci Educ Theory Pract* 2017;22(4): 951–968.
36. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach* 2009;31(8):685–695.
37. Boshuizen H. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. 1992;16(2):153–184.

38. Boshuizen HP, Schmidt HG. The development of clinical reasoning expertise. *Clin Reason Health Prof* 2008;3:113–121.
39. Erasmus University Medical Center, VirtualMedSchool. *AbcdeSIM*. Rotterdam: VirtualMedSchool; 2012.
40. Holmqvist K, Andersson R. *Eye-Tracking: A Comprehensive Guide to Methods, Paradigms and Measures*. Lund, Sweden: Lund Eye-Tracking Research Institute; 2017.
41. Tien G, Zheng B, Atkins MS. Quantifying surgeons' vigilance during laparoscopic operations using eyegaze tracking. *Stud Health Technol Inform*. 2011;163:658–662.
42. Tien G, Atkins MS, Zheng B, Swindells C, eds. Measuring situation awareness of surgeons in laparoscopic training. In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. New York: Association for Computing Machinery; 2010.
43. Aldekhyl S, Cavalcanti RB, Naismith LM. Cognitive load predicts point-of-care ultrasound simulator performance. *Perspect Med Educ* 2018;7(1):23–32.
44. Patton MQ. Enhancing the quality and credibility of qualitative analysis. *Health Serv Res* 1999;34(5 Pt 2):1189–1208.
45. Naismith LM, Cavalcanti RB. Validity of cognitive load measures in simulation-based training: a systematic review. *Acad Med* 2015;90(Suppl 11):S24–S35.
46. Paas FGWC. Training strategies for attaining transfer of problem-solving skill in statistics—a cognitive-load approach. *J Educ Psychol*. 1992;84(4):429–434.
47. Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007;41(12):1133–1139.
48. Haig KM, Sutton S, Whittington J. SBAR: a shared mental model for improving communication between clinicians. *Jt Comm J Qual Patient Saf* 2006;32(3):167–175.
49. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
50. Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;67(1):1–48.
51. Seufert T. The interplay between self-regulation in learning and cognitive load. *Educ Res Rev*. 2018;24:116–129.
52. Lei Z, Waller MJ, Hagen J, Kaplan S. Team adaptiveness in dynamic contexts. *Group Organ Manag*. 2016;41(4):491–525.
53. Rudolph JW, Repenning NP. Disaster dynamics: understanding the role of quantity in organizational collapse. *Adm Sci Q* 2002;47(1):1.
54. Waller MJ, Uitdewilligen S. Talking to the room: collective sensemaking during crisis situations. In: Roe RA, Waller MJ, Clegg SR, eds. *Time in Organizational Research*. Milton Park, Abingdon, Oxon; New York, NY: Routledge; 2008:208–225.
55. Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007;77(1):81–112.
56. Bardach L, Klassen RM, Durksen TL, Rushby JV, Bostwick KC, Sheridan L. The power of feedback and reflection: testing an online scenario-based learning intervention for student teachers. *Comput Educ* 2021;104:194.