

<https://doi.org/10.1038/s41539-025-00380-9>

Prior text-based learning enhances the benefits of simulation-based learning on confidence and long-term recall

Angélique Lebert ^{1,2} & Óscar Vilarroya ^{1,2}

Simulation-based learning (SBL) has been mostly studied in medical and STEM settings, where its specific role in learning is difficult to isolate. We ran a randomized, counterbalanced mixed design ($N = 88$) in which naïve adults learned two fictitious clinical cases via a live standardized-patient simulation or a structured text format; content and exposure time were matched. One month later, video- and text-based evaluations assessed accuracy, reaction time, and confidence. SBL yielded faster responses and higher accuracy than text. An order effect emerged: learning text first boosted subsequent SBL accuracy, while the opposite order did not. Individual imagery ability predicted better accuracy overall and more strongly in the text condition. Participants were more likely to report absolute confidence after SBL, with confidence more closely aligned to accuracy. These findings suggest that both SBL and text support durable learning, but their effects depend on instructional order and imagery ability, and SBL additionally enhances confidence-accuracy calibration.

Simulation-based learning (SBL) is an educational approach that uses realistic, interactive, and immersive simulations to enhance learning and skill acquisition^{1,2}. By replicating real-world scenarios, it offers hands-on experiences and serves as an alternative to traditional learning methods. SBL has been widely applied in specialized fields such as science, technology, engineering, and mathematics (STEM) and medical education to develop procedural skills, communication abilities, decision-making, and the application of theoretical knowledge in controlled environments^{3,4}.

Three key aspects of SBL are believed to enhance learning. First, it provides a realistic experience by immersing learners in scenarios featuring real or virtual actors, tools, and contexts that closely resemble those in which the acquired skills or knowledge will be applied⁴. Second, SBL is typically designed to create a sense of immersion, fully engaging a learner's attention and perception and making them feel deeply involved or absorbed¹. Finally, SBL introduces an embodied and interactive dimension that strengthens learning by requiring active engagement with the environment, material objects, and other individuals. The embodied aspect grounds learning in real experiences, while the interactive component dynamically reinforces cognitive processes through action and response rather than passive reception⁵.

Research on SBL has primarily been conducted in structured educational settings, such as classrooms, professional training programs, and hospital-based instruction⁶. These environments integrate SBL with relevant content, materials, and instructional methods, often incorporating complementary support such as lectures, guided reflections, worked

examples, or feedback during or after simulations, facilitating the impact of SBL interventions^{4,6,7}. While these elements likely enhance learning, their specific contributions are not always systematically controlled, making it difficult to isolate the unique role of SBL in the learning process^{4,6–11}.

Moreover, while some SBL studies employ randomized controlled trials (RCTs), practical constraints often shape sample sizes, comparison conditions, requiring a balance between ecological validity and experimental control. Consequently, comparison groups and control conditions may differ in important aspects, complicating attributions of learning differences to SBL itself. Variability of the control group^{8,9}, content exposure⁷, learning duration, instructional setting^{4,6}, assessment analyses and methods across conditions^{6,7,10,12}, potential biases from non-blinded assessments (e.g., by experts)^{7,8}, and limited assessments of long-term retention^{7,8} further complicates interpretation. Additionally, the widespread use of between-subjects designs provides insight into overall learning effects but introduces inter-individual variability.

In sum, while prior research highlights SBL's potential, methodological complexities make it challenging to assess its true impact. Variations in instructional design, engagement levels, and assessment procedures raise questions about whether observed performance differences genuinely reflect the benefits of SBL or stem from extraneous factors. Addressing these limitations through a rigorously controlled study with a robust experimental design in a lab setting is crucial for accurately determining SBL's advantages over traditional learning methods.

¹Unitat de Recerca en Neurociència Cognitiva, Departament de Psiquiatria i Medicina Legal, Universitat Autònoma de Barcelona, Barcelona, Spain. ²Hospital del Mar Research Institute, Barcelona, Spain. ✉e-mail: Oscar.Vilarroya@uab.cat

This study examines the effectiveness of an SBL intervention compared to text-based learning in acquiring and retaining novel information. Participants, who were carefully selected to have no prior training or familiarity with medical, psychological, or psychiatric conditions, learned two distinct and fictitious clinical cases with structured symptom sequences, either through interactive simulation with an actor or through a text-based method. The experimental design systematically controlled for order effects and task biases by counterbalancing conditions and symptom presentations.

To ensure that the acquired knowledge was entirely novel and uncontaminated by preexisting associations, we created two fictitious syndromes modeled on realistic but non-existent clinical conditions. This approach provided an additional safeguard against the possible use of prior knowledge, even accidental, that could arise if real disorders were employed. By combining naïve participants with entirely novel learning material, we ensured that any differences in performance reflected the learning conditions themselves rather than familiarity with clinical content.

Clinical cases, widely used in SBL, provided us with an optimal balance between acquiring specific knowledge and enabling its representation through bodily movements. The structured, sequential symptom presentation mirrors realistic clinical encounters while enhancing the representability of the learned information. In addition, observational learning has been shown to improve retention of relational structures, such as sequential item patterns^{13,14}, hence the decision to use an actor simulating the cases. Moreover, the actor not only presented the clinical case but also engaged participants through controlled physical interactions, ensuring an active learning process rather than passive observation. Research shows that touch affects memory recognition^{15,16}, gesturing during encoding enhances recall similarly to action¹⁷, and body posture facilitates autobiographical memory retrieval¹⁸. More broadly, embodied effects, such as the “enactment effect”¹⁹, have shown to enhance retention and recall²⁰. In sum, this protocol allowed us to integrate realism and embodied interactivity—key features of SBL—immersing participants in a lifelike learning experience.

While observational learning and internal mental simulation are likely to occur across many learning formats, our study operationalizes SBL more narrowly as immersive, embodied, and interactive engagement with another agent. Specifically, we use “SBL” in the established sense of a standardized-patient (SP) encounter: a trained actor portrays a clinical case and interacts reciprocally with the learner—verbally and through brief tactile engagement—under a standardized script.

Unlike text-based learning, which can elicit mental imagery but remains non-interactive, our live-actor condition involved reciprocal physical and verbal interaction, requiring participants to actively formulate questions and adapt to the actor’s responses. In addition, the actor guided participants to touch them in order to directly perceive the manifestation of symptoms, ensuring that the learning experience was not only interactive but also physically embodied.

According to the Integrated Model of Text–Picture Comprehension²¹, text and pictures engage partly distinct processing routes, descriptive versus depictive, that converge on a single mental model. Depictive input (e.g., dynamic visual or embodied cues) tends to scaffold that model more directly, whereas text affords finer control over processing. In our study, the SBL condition implements a depictive, embodied route, and the text condition targets the descriptive route, with content held constant across modalities. Prior work on sequencing further suggests that depictive/immersive formats presented before verbal/text formats can support mental-model construction in some contexts, although effects are conditional and not uniform across outcomes (e.g., stronger at the knowledge level than on deeper measures)^{22,23}.

Long-term memory was assessed using item recognition and serial recall tasks, which are widely employed to measure retrieval strength and temporal organization^{24–27}. These methods have also been used to investigate the effects of simulation, embodiment, and embedded learning²⁰. Learning outcomes were assessed using identical objective measures across both conditions, ensuring that retrieval mode did not influence results. The study also included both video-based and written evaluations to assess

whether knowledge acquired through SBL effectively transfers across different retrieval contexts.

Beyond learning and retention, we examine the impact of an imagery questionnaire on accuracy and reaction time (RT) across different retrieval modes. Specifically, we explore whether an individual’s perceived capacity for mental imagery influences retrieval accuracy, particularly in the written evaluation modality. Prior research has shown that mental imagery can substitute for direct perceptual experience by engaging neural mechanisms similar to actual perception²⁸. This internal simulation not only allows individuals with stronger imagery abilities to reconstruct learned information more accurately and efficiently in contexts lacking external cues but also strengthens recall for abstract concepts when linked to sensorimotor experiences²⁹. Consequently, internal simulation appears capable of compensating for the absence of direct interaction, thereby reinforcing structured learning.

(H1) We propose that SBL’s immersive, interactive, and embodied design enhances sensorimotor engagement and internal simulation, fostering a robust internal feedback mechanism that reinforces both prediction and its verification. Accordingly, we expect that participants will exhibit faster RTs, higher accuracy, and greater confidence in the SBL condition compared to the text-based learning condition.

(H2) Additionally, due to the mixed design, clinical case separation, content equivalence, counterbalancing of conditions, and a delayed 1-month assessment, we do not anticipate a systematic effect of learning order on performance.

(H3) Given the distinct cognitive processes involved in retrieving information from video versus text, we expect differences in RT, accuracy, and confidence between the video and the text modalities. Specifically, we hypothesize that retrieval via video may lead to faster RTs, higher accuracy, and greater confidence compared to written modality retrieval.

(H4) Given that the two clinical cases were matched for body region, movement type, written length, and difficulty, we do not expect a main effect of Clinical Case on performance.

(H5) We predict that individuals with higher imagery capacity may exhibit faster RT and accuracy, particularly in the written modality, where internal simulation may compensate for the lack of external sensory cues.

In sum, by systematically comparing SBL and text-based learning while controlling for prior knowledge and ensuring content equivalence and design control, this study contributes to a broader understanding of how embodied and interactive learning experiences shape knowledge acquisition and retention.

It is important to note that SBL encompasses a wide variety of pedagogical practices, typically involving not only realism, immersion, and embodiment, but also decision-making, feedback, and opportunities for iterative practice. In the present study, we deliberately focused on a restricted subset of these components, specifically the immersive, embodied, and interactive dimensions, in order to experimentally isolate their contribution to learning. This approach does not capture the full richness of SBL as implemented in educational and clinical training contexts, but it constitutes a necessary first step in an incremental research program. By beginning with controlled, counterbalanced conditions in a laboratory setting, we aim to establish a foundation for future studies that will progressively incorporate additional features of SBL, moving toward greater ecological validity.

Results

Effects of condition, order, modality, and imagery on reaction time

Following the approach described in the “Data processing and statistical analyses” section (Formula 1), we fit a linear mixed-effects model on log-transformed RTs to test *H1* (Condition: simulation vs. control), *H2* (Order: simulation-control vs control-simulation), *H3* (Modality: written vs. video), and *H4* (Clinical case: P1 vs P2).

We observed a significant main effect of Condition, with participants in the control condition exhibiting longer RTs than those in the simulation condition (Estimate = 0.066, 95% CI [0.03, 0.10], SE = 0.017,

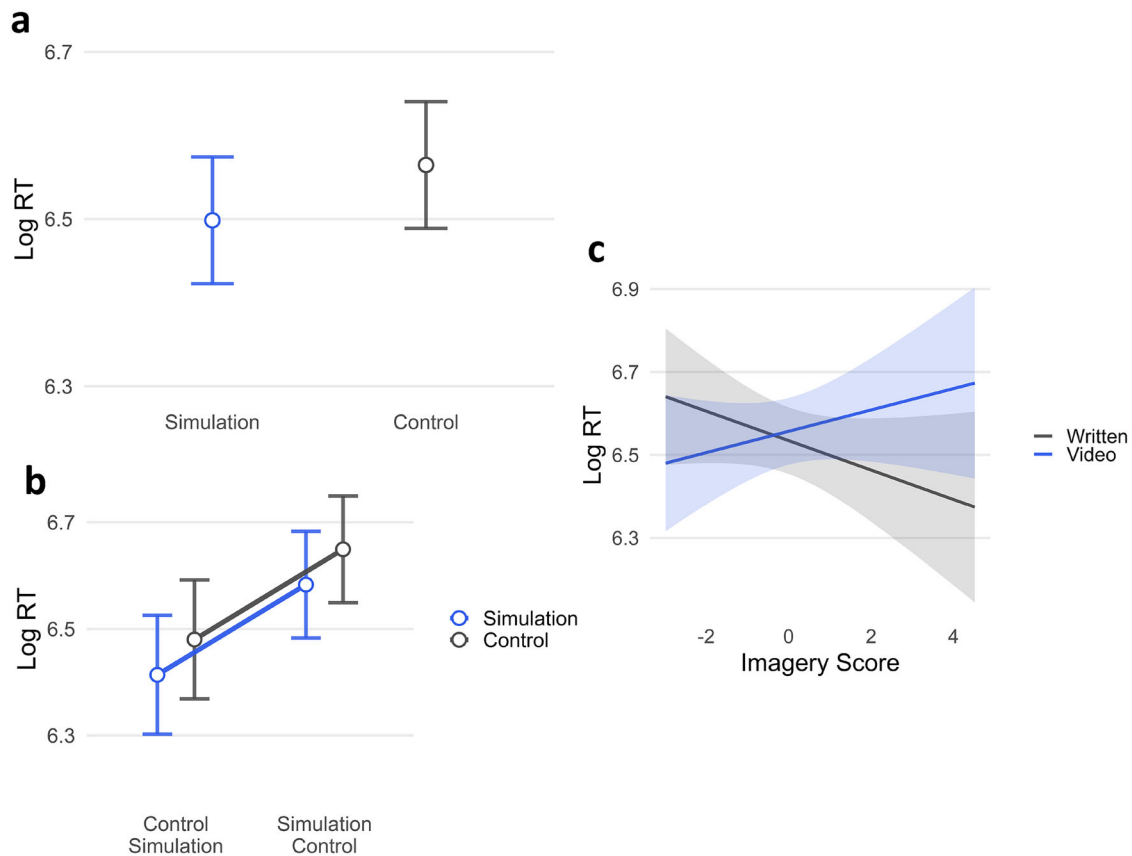


Fig. 1 | Reaction time. **a** Mean log-transformed reaction times (RT) by Condition (simulation in color vs. control in gray); error bars represent within-subject 95% confidence intervals. **b** Mean log-transformed RTs by Condition and Order (simulation-control in color vs. control-simulation in gray); error bars indicate within-subject 95% confidence intervals. **c** Predicted log-transformed RTs as a

function of participants' centered imagery score and Modality (video in color vs. written in gray), with shaded regions representing 95% confidence intervals. Estimates are derived from linear mixed-effects models including random intercepts for participants.

$t = 3.90$, $p < 0.001$). There was also a significant main effect of Order (Estimate = 0.169, 95% CI [0.02, 0.32], SE = 0.074, $t = 2.27$, $p = 0.026$), indicating that participants tested in the simulation-control sequence responded more slowly than those in the control-simulation sequence. Neither Modality (Estimate = 0.027, 95% CI [−0.01, 0.06], SE = 0.017, $t = 1.57$, $p = 0.116$) nor Clinical case (Estimate = 0.043, 95% CI [−0.10, 0.19], SE = 0.074, $t = 0.58$, $p = 0.566$) reached statistical significance. The fixed effects explained approximately 1.7% of the variance in log-RT (marginal $R^2 = 0.017$), whereas including random intercepts for participants increased the explained variance to about 24.4% (conditional $R^2 = 0.244$).

We focus here solely on the results concerning the imagery score. For the full model results, please refer to the Supplementary Table 4. Consistent with H5, we tested whether Imagery predicts RT and whether this effect differs by Modality. We did not observe a significant main effect of the imagery score on RTs (Estimate = −0.037, 95% CI [−0.08, 0.01], SE = 0.023, $t = -1.57$, $p = 0.121$), suggesting that individual differences in imagery ability alone did not influence response times. However, we observed a significant interaction between Modality and the imagery score, indicating that the effect of the modality of presentation on RTs varied as a function of participants' ability to generate mental imagery (Estimate = −0.061, 95% CI [−0.08, −0.04], SE = 0.010, $t = -5.88$, $p < 0.001$) (Fig. 1).

Effects of condition, order, modality, and imagery on accuracy

Using the generalized linear mixed-effects model described in the “Data processing and statistical analyses” section (Formula 2), we tested hypotheses H1 (Condition: simulation vs. control), H2 (Order: simulation-control

vs control-simulation), H3 (Modality: written vs. video), and H4 (Clinical case: P1 vs P2) on the likelihood of a correct response.

We observed a significant main effect of Condition, indicating that participants in the simulation condition exhibited higher accuracy than those in the control condition (Estimate = −0.456, SE = 0.097, $z = -4.71$, $p < 0.001$, corresponding to OR = 0.63, 95% CI [0.52, 0.77]). A main effect of Order also emerged, showing that participants tested in the simulation-control sequence had lower accuracy than those in the control-simulation sequence (Estimate = −0.837, SE = 0.343, $z = -2.44$, $p = 0.015$; OR = 0.43, 95% CI [0.22, 0.85]). Neither Modality (Estimate = −0.059, SE = 0.096, $z = -0.61$, $p = 0.539$, OR = 0.94, 95% CI [0.78, 1.14]) nor Clinical case (Estimate = 0.203, SE = 0.338, $z = 0.60$, $p = 0.548$, OR = 1.23, 95% CI [0.63, 2.38]) reached statistical significance. Model fit was further supported by an area under the curve (AUC) of 0.901, reflecting an excellent discriminative ability. Regarding effect sizes, the marginal R^2 was approximately 3.2%, while the conditional R^2 rose to around 55%.

We focus here solely on the results concerning the imagery score. For the full model results, please refer to the Supplementary Table 5. Consistent with H5, we tested whether Imagery predicts accuracy and whether this effect differs by Modality. Participants with higher imagery scores exhibited greater accuracy overall (Estimate = 0.266, SE = 0.114, $z = 2.34$, $p = 0.019$; OR = 1.30, 95% CI [1.04, 1.63]). Moreover, a significant interaction between imagery score and modality indicated that the relationship between imagery score and accuracy varied across modalities (Estimate = −0.186, SE = 0.058, $z = -3.22$, $p = 0.001$; OR = 0.83, 95% CI [0.74, 0.93]). Specifically, the positive effect of imagery score on accuracy was more pronounced in the written condition than in the video condition (Fig. 2).

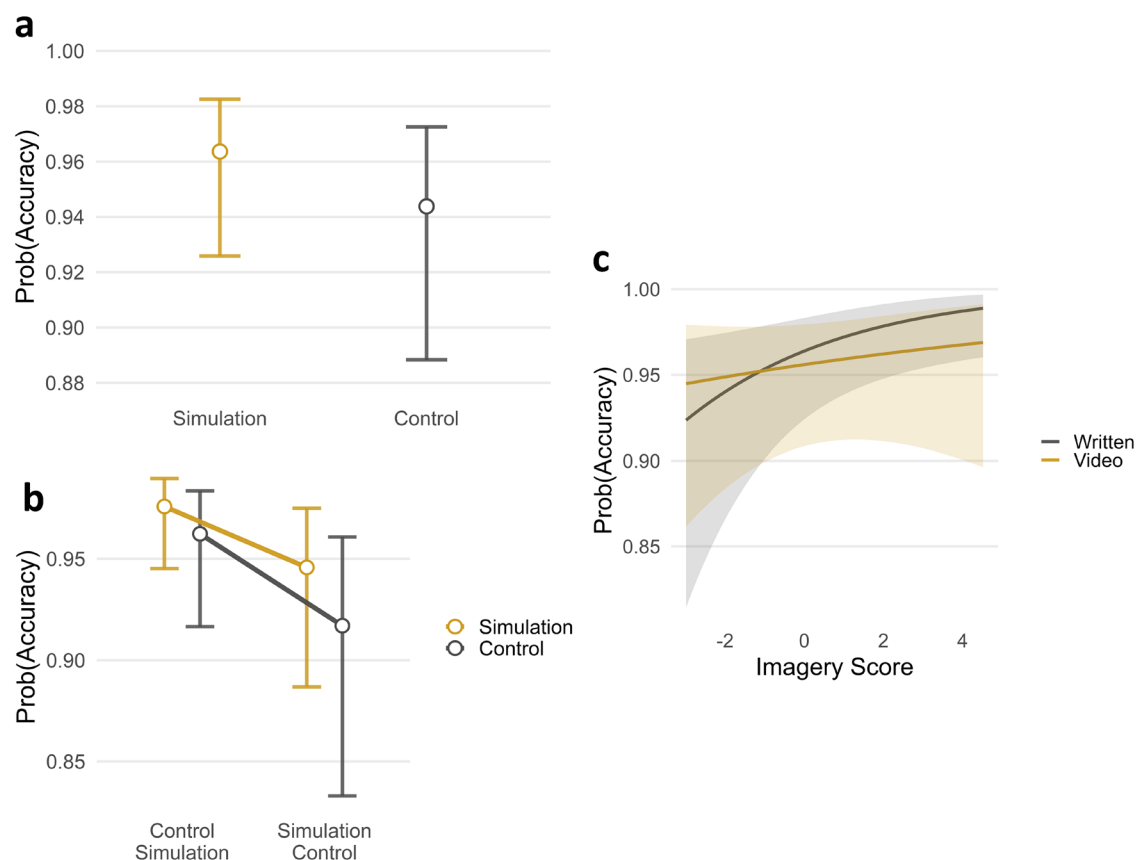


Fig. 2 | Accuracy. **a** Predicted probability of correct responses (accuracy) by Condition (simulation in color vs. control in gray); error bars represent within-subject 95% confidence intervals. **b** Predicted accuracy by Condition and Order (simulation-control in color vs. control-simulation in gray); error bars indicate within-subject 95% confidence intervals. **c** Predicted accuracy as a function of participants'

centered imagery score and Modality (video in color vs. written in gray), with shaded regions representing 95% confidence intervals. Estimates are derived from generalized linear mixed-effects models including random intercepts for participants and for stimuli.

Effects of condition, order, and modality on confidence

Using the generalized linear mixed-effects model described in the “Data processing and statistical analyses” section (Formula 3), we tested the confidence component on *H1* (Condition: simulation vs. control), *H2* (Order: simulation-control vs. control-simulation), *H3* (Modality: written vs. video), and *H4* (Clinical case: P1 vs P2).

A significant main effect of Condition emerged, indicating that participants in the control condition were substantially less likely to report absolute confidence compared to those in the simulation condition (Estimate = -1.14 , SE = 0.08 , $z = -13.68$, $p < 0.001$; OR = 0.32 , 95% CI [0.27 , 0.38]). Modality showed a marginal effect, with participants in the video condition being slightly more likely to report the highest certainty compared to those in the written condition (Estimate = 0.15 , SE = 0.08 , $z = 1.90$, $p = 0.058$; OR = 1.16 , 95% CI [1.00 , 1.36]). Neither Order (Estimate = -0.35 , SE = 0.57 , $z = -0.61$, $p = 0.541$; OR = 0.70 , 95% CI [0.23 , 2.17]) nor Clinical case (Estimate = -0.36 , SE = 0.57 , $z = -0.63$, $p = 0.527$; OR = 0.70 , 95% CI [0.23 , 2.13]) reached statistical significance. The model showed excellent predictive accuracy, as reflected by an AUC of 0.91 , indicating a high ability to discriminate between confidence levels. Regarding effect sizes, the marginal R^2 was approximately 3.5% , whereas the conditional R^2 rose to about 71.1% (Fig. 3).

Complete model outputs, including diagnostic plots and additional details, can be visualized by running the analysis scripts available in the OSF repository.

Discussion

This study, designed as an RCT using a mixed design, examined how a simulation-based protocol impacted the acquisition and long-term

retention of new knowledge by naïve learners. A notable feature of this study lies in its carefully controlled design and counterbalanced structure, allowing for reliable comparisons between learning conditions. By doing so, we were able to isolate the specific contribution of simulation-based, immersive, and embodied interaction to long-term learning outcomes.

Participants engaged with an actor who portrayed a clinical case, experiencing the symptoms dynamically through physical interaction, while a matched text-based learning condition provided identical information in a non-interactive, passive format. One month after learning, participants exhibited significantly faster RTs and higher accuracy when recalling clinical symptom sequences in the simulation-based condition compared to the text-based condition.

Interestingly, an order effect emerged: when participants began with the control condition, their subsequent performance in the simulation condition was enhanced, whereas starting with the simulation condition did not provide the same benefit for the control condition. This may suggest a potential beneficial transfer effect that enhances performance in the simulation condition when it follows the text-based condition, while the control condition undergoes a disadvantageous transfer effect when it follows the simulation condition.

Beyond these effects, we also observed that participants were significantly more likely to report absolute confidence in the SBL condition. Moreover, confidence and accuracy were closely aligned in the SBL condition, with higher confidence accompanying higher accuracy. This contrasts with the control condition, where accuracy was relatively high, but the probability of absolute confidence remained moderate, potentially leading to a more tentative application of knowledge.

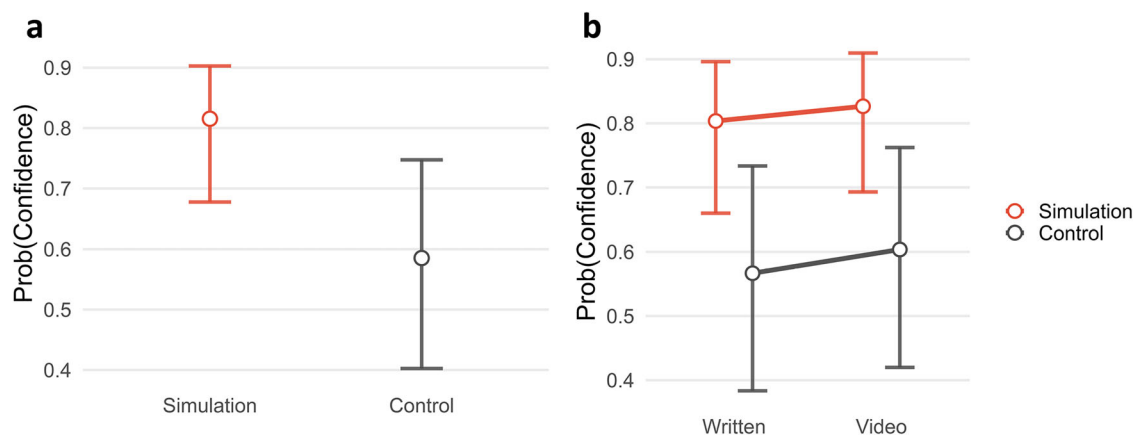


Fig. 3 | Confidence outcome. a, b Results for Confidence. **a** Mean confidence ratings by Condition (simulation case in color vs. control case in gray); vertical bars indicate within-subject 95% confidence intervals. **b** Mean confidence ratings by Condition

(simulation case in color vs. control case in gray) and Modality; vertical bars indicate within-subject 95% confidence intervals.

Our findings support previous research demonstrating the benefits of SBL in medical and STEM education^{3,4}, although they moderate the advantages of SBL over more traditional text-based methods in learning outcomes. Consistent with prior work, SBL appears to improve knowledge retention, RT, and retrieval accuracy, supporting the possibility that such experiences may contribute to stronger memory traces than passive learning methods^{30,31}. Furthermore, research emphasizing the role of embodiment and interactivity in cognitive processes⁵ suggests that learning through active engagement with a simulated patient reinforces memory processes, especially by leveraging sensorimotor integration, a mechanism shown to facilitate encoding and retrieval^{32,33}.

However, our study also provides a more nuanced view of SBL's effectiveness by isolating its specific contributions through a counter-balanced mixed design. While SBL demonstrated an advantage, text-based learning can still effectively encode and retrieve information, at least in the domain of knowledge acquisition, where information is systematically organized and sequentially presented. The observed differences between conditions might nevertheless have been moderated by the role of mental imagery and the assessment modality. On the one hand, the simulation condition might have been disadvantaged by the retrieval assessment being presented in a video format, attenuating the immersive and embodied benefits of the original learning experience. This aligns with prior research suggesting that the format of retrieval assessments can impact memory performance, particularly when there is a mismatch between learning and retrieval conditions¹². Conversely, participants with higher imagery capacity may have been able to mentally simulate the clinical case in the written evaluation modality, facilitating their retrieval process. Mental imagery has been shown to engage neural mechanisms akin to actual perception, supporting perceptual learning without direct sensory input²⁸. Additionally, it strengthens recall for abstract concepts when linked to sensorimotor experiences²⁹. This suggests that internal simulation can compensate for the absence of direct interaction.

An additional notable observation of this study is the effect of condition order. For SBL, an initial exposure to the material in a structured, traditional format may facilitate later embodied learning. Although prior reviews and VR studies often report advantages for pictorial or VR first, typically confined to knowledge-level outcomes^{22,23}, our findings show the opposite: in content-matched, temporally spaced sessions with distinct cases, beginning with text improved subsequent SBL. This pattern suggests that initial symbolic structuring can scaffold later embodied encoding and underscores that sequencing effects are conditional. This finding aligns with previous research^{34–36} suggesting that SBL benefits from prior contextualization or introduction, which may help structure the learning experience before engagement in an immersive and interactive environment. Conversely, for

text-based learning, experiencing an immersive, interactive session beforehand may reduce engagement or disrupt retention, possibly because the shift from an active, sensory-rich experience to a more passive format is less conducive to sustained cognitive engagement. This aligns with prior research indicating that immersive, embodied learning can lead to deeper but more rigid knowledge acquisition, potentially hindering subsequent learning through a more abstract and static format, such as text-based instruction^{37,38}. Immersive learning also places high demands on cognitive resources, engaging sensorimotor, emotional, and social processes. From a cognitive load perspective, this intense engagement early in learning may create a form of anchoring or cognitive rigidity, making it harder to adapt to less interactive and more abstract formats³⁷, mirroring the proactive interference observed by Saylık et al.³⁸. Consequently, when working memory processing is hindered by cognitive overload, the consolidation, retrieval, and transfer of information become significantly more challenging, highlighting the advantage of a sequential, low-to-high complexity approach that progressively builds learners' long-term memory stores³⁶. While this remains speculative, future research should explore whether such order effects are driven by differences in attentional or cognitive load across modalities, from proactive interference, or shifts in engagement²⁰ across learning modalities.

We also observed that participants in the simulation condition were three times more likely to report absolute confidence than those in the control condition. One possible explanation may stem from the immersive nature of SBL, which fosters deeper sensorimotor engagement and internal simulation. As suggested by the sensorimotor model of memory, this process may facilitate the reactivation of sensorimotor patterns associated with the learned material. Specifically, the richer sensory and contextual cues provided by SBL, compared to the control condition, may act as markers of a more robust internal feedback mechanism, strengthening the link between knowledge verification and retrieval²⁰.

In addition, in our results, confidence and accuracy were closely aligned in the simulation condition, with higher confidence accompanying higher accuracy. This alignment suggests that SBL not only enhances knowledge retention but also reinforces learners' trust in their acquired knowledge. In contrast, the control condition exhibited a mismatch, where accuracy was relatively high, but absolute confidence probability remained moderate. This discrepancy may indicate a more tentative application of knowledge, as learners might hesitate despite correctly recalling information. While confidence is generally beneficial^{39–41}, its relationship with accuracy is complex⁴². When confidence and competence are well-calibrated, learners are more likely to apply their knowledge effectively³⁹. However, when confidence outpaces accuracy, overconfidence can lead to errors and misjudgments^{42,43}. Therefore, SBL was associated with both

higher accuracy and greater confidence, suggesting that SBL may support environments where learners both acquire and apply knowledge effectively than with more traditional methods of learning. This combination of accuracy and confidence may be particularly valuable in settings where both competence and the conviction to act on one's knowledge are essential. Moreover, by aligning confidence with performance, SBL may help mitigate the risks of overconfidence.

Future studies should aim to replicate these findings and further investigate the mechanisms underlying the observed alignment between high absolute confidence probability and accuracy in SBL. Specifically, research should explore how sensorimotor engagement and internal simulation contribute to learners' self-assessment and whether the enhanced sensory and contextual cues in SBL function as an internal feedback mechanism that strengthens learning outcomes. Additionally, examining the conditions under which confidence and accuracy remain well-calibrated in SBL, as well as identifying potential boundary effects, would provide deeper insights into the robustness and generalizability of these effects.

Interestingly, our results indicate that despite the advantages of SBL, text-based learning also appeared to facilitate item identification and sequential recall of realistic, embodied knowledge, enabling learners to transfer this information to recognition tasks involving realistic symptom presentations. Related to this, our findings suggest a possible role of mental imagery in knowledge retrieval when confronted with its realistic depiction. Participants with stronger imagery abilities performed better, suggesting that the ability to mentally simulate content may be an important predictor of retrieval efficiency. This finding extends beyond SBL and may have implications for understanding individual differences in learning across various modalities, such as virtual reality training, medical case studies, or language learning, where visualization plays a key role in comprehension and recall.

Regarding limitations, although our design was sufficiently powered for main effects, it was less sensitive to detect higher-order interactions, which should therefore be interpreted with caution. In addition, despite the fact that the study demonstrated an advantage of SBL over text-based learning within a 1-month retention period, future studies should examine whether the observed benefits persist beyond 1 month and whether different retrieval modalities influence retention over extended periods.

Another limitation concerns the nature of the retrieval assessments. Since knowledge retention was tested using video assessments rather than direct interactions with a live actor, the transferability of learning from SBL to real-world settings may have been attenuated. It is possible that a more immersive retrieval condition, such as role-playing with a live patient, would have amplified the benefits of SBL by preserving its embodied and interactive elements.

In addition, the order effect, while intriguing, requires further validation. The fact that simulation benefited from an initial text-based session while text-based learning was negatively impacted by starting with simulation suggests that cognitive flexibility, engagement shifts, or proactive interference may play a role. Future studies should explore whether this effect generalizes to other educational contexts and whether instructional sequencing can be strategically optimized to enhance learning outcomes.

More generally, a strength, but also a weakness, of this study is that it specifically targeted the realistic, immersive, embodied, and interactive dimensions of SBL in a controlled laboratory setting. Accordingly, our findings pertain to SP-based SBL—a well-established subtype characterized by reciprocal verbal and tactile interaction and social presence—rather than to SBL writ large. Because our SP condition bundles several properties (embodiment, interactivity, social richness) while the text-plus-narration control was deliberately non-interactive and abstract, the observed advantages should be interpreted as the effect of this embodied-interactive bundle relative to text, not as the isolated effect of “simulation” in the broadest sense. However, SBL also encompasses other crucial components, such as strategic decision-making, dynamic feedback, and iterative practice⁴⁴, which contribute significantly to its pedagogical value^{45–47}. While our protocol did not

provide dynamic, performance-based feedback, it did integrate reflective opportunities. After each session, participants completed questionnaires on their affective and anxiety states, encouraging reflection on their internal responses to the simulation. Moreover, after the second session, a baseline evaluation required participants to revisit and organize the symptom sequences, offering an indirect form of content-related reflection. These elements, while not equivalent to the adaptive feedback and iterative practice typically found in applied SBL, nevertheless align with experiential learning frameworks by prompting learners to process both their internal states and the material they have just encountered. We therefore consider our protocol an incremental step: by isolating the embodied and interactive aspects of SBL while embedding indirect reflective elements, we aim to contribute a methodological step for future studies that will integrate additional ecological features and capture the full pedagogical richness of SBL.

In relation to this point, a further limitation is that we did not explicitly frame our study within cognitive load theory (CLT)^{48–50}, which is a framework often used to explain how different types of learning tasks place demands on learners' working memory and, in turn, affect their ability to acquire and retain knowledge. When designing our study, we were careful to ensure that the basic difficulty of the task was equivalent across conditions. Both conditions conveyed the same symptom content with a matched structure, and participants were naïve to the domain and randomly assigned. In this sense, we aimed to ensure that content and prior knowledge were comparable across conditions. Where they diverge is in how the material was delivered and how learners were required to engage with it. To examine this more precisely through the lens of CLT, future studies will need to be explicitly designed to measure these factors directly.

In conclusion, this study offers evidence suggesting potential benefits of SBL over traditional text-based learning for the acquisition and long-term retention of clinical knowledge. While SBL was associated with enhanced learning, our results suggest that structured, text-based methods can also yield strong retention outcomes, particularly for item identification and sequential recall.

Although our design does not capture the usual SBL protocol, it provides a controlled context to isolate the contribution of realism, embodiment, and interactivity in SBL, thereby contributing to a more precise understanding of the mechanisms that may underlie its effects. Identifying potential factors that enhance learning outcomes may help inform learning protocols, whether simulation-based or not, to incorporate principles of interactivity and embodiment, maximizing their effectiveness.

This refined understanding may inform better-informed design of SBL interventions and offers insights that can be transferred to non-SBL learning environments. Future research should extend this work to domain experts and more ecologically valid scenarios, helping to develop more tailored, evidence-based approaches to enhance learning through immersive and interactive experiences.

Methods

The presented study was not preregistered. This research was approved by the Ethics Committee at the Hospital del Mar Research Institute (Ref: 2022/10237/I). All participants signed a consent form before participating in the study and received monetary compensation for their time.

Participants

Participants ranged in age from 18 to 35 years and were enrolled under pre-specified exclusion criteria: individuals with academic training or work experience in medicine or psychology, and those with any personal psychiatric, neurological, motor, or physical history, were excluded, thereby minimizing prior content knowledge (see Supplementary Note 1 for the full inclusion/exclusion criteria). They were recruited through convenience sampling by posting flyers on several university campuses.

Sample size and power analysis

A power analysis was conducted using G*Power to determine the required sample size, ensuring adequate power to detect anticipated effects

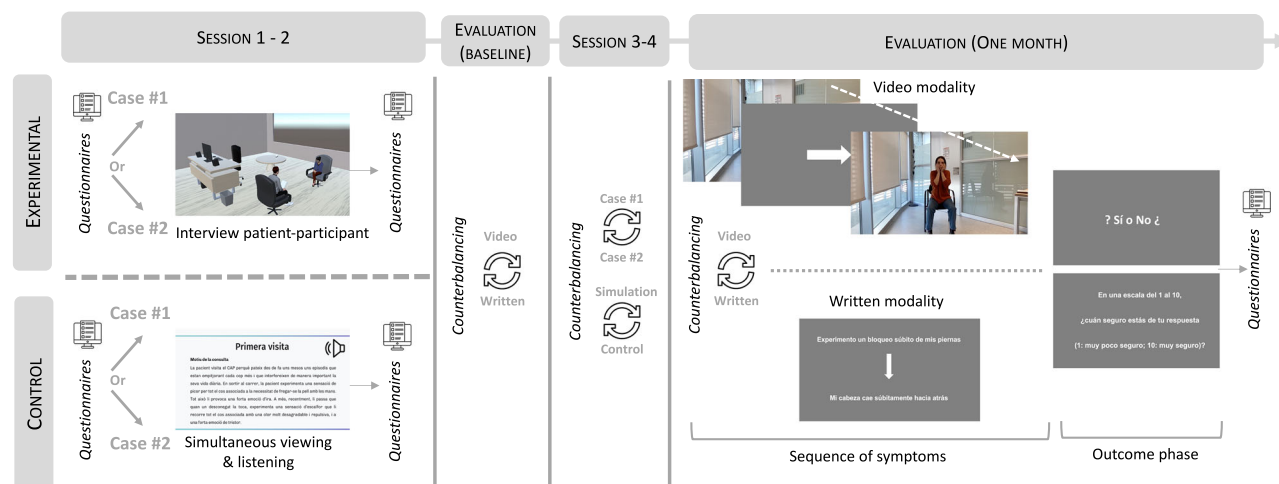


Fig. 4 | Schematic representation of the experimental procedure. This figure illustrates the arrangement of the simulation room (Unity environment, <http://www.unity3d.com>), alongside the presentation of clinical cases in the control condition across the two learning sessions. During each evaluation phase, participants were

asked to judge whether a displayed symptom sequence (video or written format) was correct and to rate their confidence in their response. Both conditions and clinical cases were systematically counterbalanced across participants to ensure balanced exposure.

(Supplementary Note 2). Ultimately, 88 participants ($n = 58$ females, 25 males, 5 non-binary, mean age $\pm$ SD = 22.98 ± 4.50) were included in the final analysis after accounting for exclusions. One participant was excluded due to a software issue, and 12 participants were excluded following the actors' video quality check (see the "Quality check of videos of the simulation sessions" section for more details).

Equipment and software

All online questionnaires were administered via the Qualtrics platform, and evaluations were conducted using OpenSesame (version 4.0⁶¹) on a Windows computer with a 23-inch Dell monitor (1920×1200 pixels), along with a keyboard and mouse. To facilitate manual responses, two stickers were placed on the keyboard, clearly indicating the designated keys for answering.

SBL room setup

A single room was used for both sessions, ensuring consistency in the learning environment. The chairs for the participant and the actor were positioned one meter apart—marked on the floor with tape—aligning with typical social spacing norms^{52,53}. A clock was placed on a table visible to both the participant and the actor, helping them manage the timing of each simulation session. Finally, two 360° pivot cameras (AOSU - C2E) were installed to record both the actor and the participant. This setup enabled real-time monitoring by the experimenter and allowed for later review if needed, thereby ensuring the simulation sessions were conducted as intended.

Simulation actors

Professional actors were recruited for both the pilot and main studies. These actors regularly collaborate with Hospital del Mar to conduct SBL sessions for medical students. Over the course of several weeks, they participated in monitored and recorded training sessions, alongside regular meetings with the study's supervisors. The primary objectives, before starting the experiment, were to ensure:

- Mastery of the clinical cases and associated symptoms,
- Consistent intensity and portrayal of symptoms across all actors,
- Minimal improvisation—restricted to prearranged anecdotes about how the symptoms affect the patient's daily life.

Stimuli

To refine the clinical cases, determine the number of sessions, and calibrate the complexity of the evaluation task, a pilot study was conducted prior to the main study (see Supplementary Note 3).

Under the guidance of an external psychiatrist (Dr. Daniel Bergé), two distinct clinical cases were developed to ensure that the acquired knowledge was entirely novel and free from preexisting biases; each case featured a fictitious syndrome characterized by three core symptoms. These symptoms were designed to involve the same bodily part and a similar bodily movement. Each case began with a triggering situation, followed by the sequential presentation of Symptom 1, Symptom 2, and Symptom 3. Additionally, each symptom was paired with a corresponding control symptom. To ensure consistency, the level of difficulty and the written length of the symptoms were standardized across the two clinical cases, maintaining a uniform level of complexity.

Evaluation materials

To evaluate symptom recall, all possible pairs were created, including combinations of symptoms, control symptoms, and mixed pairs (symptom + control symptom), resulting in 15 total combinations. Of these, only two sequences were correct (Symptom 1 $\rightarrow$ Symptom 2 and Symptom 2 $\rightarrow$ Symptom 3). Participants underwent evaluations in two distinct modalities:

In the written modality, each combination displayed the first symptom above a downward arrow, and the second symptom below. In the video modality, all symptoms and control symptoms were recorded with a different actor—one who did not take part in the SBL sessions—so that participants would not recognize them. Each video was standardized for elements such as attire, lighting, camera angle, and duration. During the combined video sequence, the first symptom was shown, followed by an arrow, and then the second symptom.

Self-report measures

Previous research has shown that social interactions can be influenced by changes in depression⁵⁴, anxiety⁵⁵, and emotional regulation strategies⁵⁶. To account for these variables in our sample, participants completed several self-report measures, including the World Health Organisation–Five Well-Being Index (WHO-5), assessing general well-being⁵⁷, the State Trait Anxiety Scales (STAI), evaluating anxiety⁵⁸, and the Scale of Positive and Negative Experience (SPANE), measuring affect⁵⁹.

Additional questionnaires assessed socio-demographic factors, potential mental health diagnoses, medication use, and current stress levels (0–10 scale). However, because analyzing how these traits might influence behavior lies beyond our scope, we provide only descriptive statistics and correlations for these measures in Supplementary Tables 1 and 2. Finally, given that previous research has shown that higher levels of embodied mental imagery can enhance learning outcomes²⁹, we included a brief

scenario-based measure of imagery, which asked participants to rate their ability to imagine key sensorimotor features (e.g., appearance, movement, voice) of a nonclinical figure and perceived physical contact with the patient in the clinical case. These ratings were combined into a single 10-point index reflecting the overall richness of participants' mental simulation (Fig. 4).

Overall experimental design

This study employed a mixed design, incorporating both within-subjects and between-subjects factors. Participants were randomly assigned to one of two groups, ensuring systematic counterbalancing of learning conditions (SBL vs. text-based) and clinical cases. Two fictitious syndromes were developed in consultation with an external psychiatrist, each characterized by three core symptoms affecting the same bodily region and involving similar motor components. Each case followed a structured sequence: an initial triggering situation, followed by the progressive presentation of Symptom 1, Symptom 2, and Symptom 3. To assess symptom recall, participants completed an evaluation phase in two distinct modalities: written and video-based. The written modality presented symptom pairs in a structured format, while the video modality displayed pre-recorded symptom demonstrations by an unfamiliar actor. Each evaluation required participants to judge whether presented symptom sequences were correct, reinforcing the study's focus on both item identification and sequential recall. To control for learning order effects, both symptom sequences and assessment modalities were counterbalanced across participants, ensuring that no systematic bias influenced the results.

Participants were randomly assigned to one of two groups, and the simulation/control conditions as well as the two cases were systematically counterbalanced to minimize order effects and any task-specific biases. Within each modality (written or video), symptom-combination presentations were randomized, and the sequence of modalities themselves was counterbalanced across participants. To indicate whether a displayed combination was correct, participants pressed one of two marked keys (left or right). Moreover, the mapping of these keys (left/right = correct/incorrect) was reversed for half of the participants to avoid response biases. To ensure further balance, half of the participants completed the simulation condition during the first week and the control condition during the second week, while the remaining half did the opposite. Within each subgroup, half of the participants were assigned to case 1 initially and then switched to case 2 for the following week (and condition), whereas the other half followed the reverse order. This design guaranteed that both the conditions and pathologies were evenly distributed among participants.

On the day of the experiment, each participant provided written informed consent upon arrival. They subsequently completed a series of questionnaires, including a socio-demographic questionnaire, the WHO-5, ERQ, STAI-Y (trait), stress assessments, and the imagery representation questionnaire, all administered via the Qualtrics platform.

In the SBL condition, participants were given a plastic sheet containing guiding questions (see Supplementary Note 4) designed to help them gather essential clinical information about the patient's symptoms. They were instructed to ask these questions during the interview, with the flexibility to reformulate them as needed. This structure ensured that participants were not passive observers but active decision-makers, choosing how to engage with the actor, when to pose questions, and how to follow up on the actor's responses. Moreover, the actor was instructed to involve participants in direct physical interaction, for example, by guiding them to touch their body to feel the symptom being manifested, thereby reinforcing the embodied dimension of the simulation. To ensure greater fluency, participants were given approximately 10 min before the start of the experiment to familiarize themselves with the full set of questions. Additionally, they were informed that the patient did not display any severe or violent symptoms.

Prior to entering the room (where the patient was already seated), the experimenter instructed participants to identify and memorize the patient's symptom sequence through an interview. They were informed that the patient might engage in tactile contact and that the sessions would be recorded; however, they were advised not to focus on the camera. Each

session had to last between 10 and 12 min, and around the 12-min mark, the experimenter would knock on the door to signal the end of the session, while also monitoring via live video to confirm that the actor had finished describing the symptoms. Participants then completed STAI-Y (state) and SPANE questionnaires. Twenty-four hours later, the same procedure was repeated with the same actor to reinforce and deepen the information. After this second session, participants again completed the STAI-Y (state) and SPANE questionnaires, followed by the case evaluation in two counterbalanced formats: video and written.

In the control condition, participants were tasked with identifying and memorizing a sequence of symptoms from a patient described in a text displayed on the screen. The video accompanying the text described one of the two clinical cases and mirrored the structure and information provided during the SBL sessions. The text appeared progressively on the screen while being narrated by an unfamiliar voice, ensuring the viewing duration was standardized at 10 min, comparable to the SBL sessions. Participants were instructed to remain focused throughout the presentation, without pausing or interruptions. Following this session, they completed the STAI-Y (anxiety) and SPANE (affect) questionnaires via the Qualtrics platform.

Twenty-four hours later, participants took part in a second session that presented more details and information about the same clinical case, emphasizing the symptoms and their impact on the patient's everyday life. After finishing this second session, participants once again filled out the anxiety-state and affect questionnaires and then carried out the evaluation of the pathology in both video and written modalities.

Information equivalence across modalities

To ensure that both modalities delivered the same case content, each case was governed by a single script that fixed the content and order of symptom information and followed a standardized clinical interview structure (e.g., introduction, personal information, reason for consultation, illness history, daily routines, etc) (see Supplementary Note 4). The control condition recast the same script as a narrated, progressively displayed text with matched exposure time, and the second session in each modality introduced the same additional details. While minor delivery differences are inherent to distinct modalities, case content and sequencing were held constant across conditions.

Initial baseline assessment and 1-month evaluation

To ensure that participants engaged with the material comparably across conditions before progressing to the long-term retention phase, we conducted an initial baseline assessment immediately after the learning phase. This assessment served as a methodological check to confirm that there were no systematic differences between conditions, clinical cases, or counterbalancing factors.

During the initial baseline assessment, participants were seated at a viewing distance of approximately 60 cm from the screen, in line with recommendations for comfortable vision⁶⁰, with audio levels standardized across all participants. The experimenter provided verbal instructions, informing participants that they would view either written symptoms (written-modality evaluation) or short video clips (video-modality evaluation) on the screen. Participants were instructed to pay close attention to the order in which the symptoms appeared. Following this, participants read detailed on-screen instructions explaining that they would need to determine whether the displayed symptom sequence corresponded to the patient they had encountered during the previous two sessions (SBL or control condition).

Each trial began with a fixation cross displayed at the center of the screen for 2 s. The symptom sequence was then presented for 10 s in the written modality or for 15–20 s in the video modality. After the sequence disappeared, a response screen prompted participants to indicate as quickly and accurately as possible whether the sequence was correct by pressing the designated key. Participants were then asked to rate their confidence in their response on a scale from 1 to 10. A 3-s interval separated each trial. The order of evaluations was counterbalanced, with some participants completing the written-modality evaluation first, followed by the video-modality

evaluation, and others completing the evaluations in the reverse order. Measures reflecting performance speed (i.e., response time between the display of the answer screen and the keypress), accuracy (i.e., correct or incorrect keypress), and participants' self-reported confidence (rated using the mouse) were recorded for each trial.

One-month session

One month after the last session, participants performed an evaluation in both the video and written formats. The key difference in this session was that participants were required to identify which correct symptom sequences they had observed in any of the previous sessions (SBL and control combined). They were instructed to respond both rapidly and accurately, and to remain focused, as they could not pause the trial or revise their answers. After completing this evaluation, participants filled out the STAI-Y (state) and SPANE questionnaires for the final time.

Quality check of videos of the simulation sessions

Whenever there was uncertainty about a given session's adherence to instructions, we conducted a quality check of the video recordings to identify any potential issues related to noncompliance with the experimental protocol. For instance, some exclusions were due to the actor not performing the interaction correctly with the participant, such as not properly engaging in the sensorimotor interaction.

Data processing and statistical analyses

All inferential tests were performed using RStudio (version 4.1.2⁶¹). Using Tukey's method to detect extreme outliers ($\pm 3 \times \text{IQR}$), we replaced outlier trials with the overall mean RT calculated without including outliers. In the baseline evaluation session, 2.53% of the trials were replaced, while in the 1-month evaluation session, 5.26% were replaced using this method^{62,63}.

All significance tests were two-sided, considered significant when $p < 0.05$. For repeated-measures analyses, if the sphericity assumption was violated, Greenhouse–Geisser corrections were applied. Effect sizes are reported as partial eta-squared (or eta-squared, as appropriate).

We began by assessing whether demographic variables and counterbalancing factors influenced our primary outcomes in our initial Baseline Assessment. Using the *aov* function (*afex* library), we examined the effects of gender, actor portraying the patient, lateralization of correct/incorrect responses on the keyboard, and written vs. video evaluation order on RT and accuracy. These analyses revealed no significant effects. We also confirmed that neither the Order of conditions (simulation first vs. control first), the Condition (simulation vs. control), nor the Clinical case (P1 vs. P2) had any observable impact on these measures. Detailed statistical outputs from this baseline assessment are provided in the Supplementary Table 3.

Subsequently, our main interest lay in examining performance 1 month later, focusing on RT, accuracy, and confidence. Because raw RT did not meet the normality assumption and were skewed to the right, we applied a logarithmic transformation. This improved residual normality and homoscedasticity, as confirmed by Q-Q and Scale–Location diagnostics (*easystats* in R); diagnostic outputs are provided in the Supplementary Information and OSF repository.

We employed linear mixed-effects models using *lmer* (*lmerTest* package) and generalized linear mixed-effects models using *glmer* (*lme4* package) where appropriate. Fixed effects were pre-specified on theoretical grounds—Condition and Evaluation modality as primary predictors; Order (sequence) and Clinical case as control variables—and no data-driven model selection was conducted.

For RT, we fit an LME model with log-transformed RT as the dependent variable and Condition (simulation vs. control) and Modality (written vs. video) as fixed effects. We also included the Order (simulation-control vs control-simulation) and Clinical cases (P1 vs P2) as control variables to account for counterbalancing and case-specific differences. Moreover, we incorporated a random intercept to control for subject-specific differences (Formula 1). The model converged successfully, and diagnostic checks (e.g., residual normality, homogeneity of variance, and

collinearity) indicated a satisfactory model fit.

$$\text{Reaction Time} \sim \text{Condition} + \text{Modality} + \text{Order} + \text{Clinical case} + (1|\text{ID}) \quad (1)$$

Accuracy (0 = incorrect, 1 = correct) was analyzed via a generalized linear mixed-effects model (binomial link). We used Condition and Modality as fixed effects, and we also included the Order of conditions and the Clinical case as control variables. Lastly, to account for subject- and stimulus-specific differences, we incorporated random intercepts into the models (Formula 2). The model converged successfully, diagnostics indicated no issues with overdispersion or zero-inflation, and outlier tests did not reveal any problematic observations.

$$\text{Accuracy} \sim \text{Condition} + \text{Modality} + \text{Order} + \text{Clinical case} + (1|\text{ID}) + (1|\text{STIM}) \quad (2)$$

Confidence was categorized into absolute (100%) certainty versus anything below 100%. Separating them thus allowed us to capture potential qualitative differences between those who were entirely certain and those who expressed any level of uncertainty, reflecting a distinctly different cognitive state. Given the large cluster of self-reported scores at 100%, this approach also balanced the data for logistic regression. To ensure the reliability of our logistic regression model and in line with the events per variable (EPV) recommendations^{64,65}, we confirmed that our model maintained an adequate EPV and that predictor variables were well-balanced, eliminating the need for penalized likelihood corrections (for further details, see the Supplementary Note 5). Hence, we fitted GLME model using confidence as the dependent variable and Condition as well as Modality as fixed effects. We included Order and Clinical case as control variables, and we incorporated random intercepts to control for subject and stimuli specific differences (Formula 3). Diagnostic checks showed no indications of singularity, over/under-dispersion, zero-inflation, or problematic outliers.

$$\text{Confidence} \sim \text{Condition} + \text{Modality} + \text{Order} + \text{Clinical case} + (1|\text{ID}) + (1|\text{STIM}) \quad (3)$$

We employed the *DHARMA* package to evaluate model adequacy in both accuracy and confidence analyses. Residual plots, dispersion tests, checks for zero-inflation, and outlier detection indicated no major violations of model assumptions. We further verified our classification approach by examining the ROC curve and computing the AUC for accuracy and confidence, confirming the appropriateness of these binary outcomes. Complete statistical outputs, diagnostic plots, and the corresponding R code are provided in the Supplementary Information.

Finally, we examined the effect of the imagery score on RT and accuracy during the 1-month evaluation. Due to missing responses, this analysis was conducted on a sample of 82 participants ($n = 56$ females, 22 males, 4 non-binaries; mean age $\pm$ SD = 22.96 ± 4.52). The imagery score was standardized (mean-centered and scaled) to facilitate interpretation and to ensure comparability across participants.

We first fit a similar LME model using log-transformed RT as the dependent variable and an interaction between Modality and the Imagery score as fixed effects (Formula 4):

$$\text{Reaction Time} \sim \text{Condition} + \text{Modality} * \text{Imagery score} + \text{Order} + \text{Clinical case} + (1|\text{ID}) \quad (4)$$

Then, we fit a similar GLME model with accuracy as the dependent variable and an interaction between Modality et the imagery score as fixed factors (Formula 5):

$$\text{Accuracy} \sim \text{Condition} + \text{Modality} * \text{Imagery score} + \text{Order} + \text{Clinical case} + (1|\text{ID}) + (1|\text{STIM}) \quad (5)$$

Data availability

All data are available at the following OSF repository: <https://osf.io/ub6ce>.

Code availability

Data preprocessing and analysis scripts are available at the following OSF repository: <https://osf.io/ub6ce>.

Received: 18 March 2025; Accepted: 6 November 2025;

Published online: 24 November 2025

References

- Frasson, C. & Blanchard, E. G. Simulation-based learning. In *Encyclopedia of the Sciences of Learning* (ed. Seel, N. M.) 3076–3080. https://doi.org/10.1007/978-1-4419-1428-6_129 (Springer US, 2012).
- Gormley, G. J. & Murphy, P. When I say ... simulation. *Med. Educ.* **57**, 1182–1183 (2023).
- Chernikova, O. et al. Simulation-based learning in higher education: a meta-analysis. *Rev. Educ. Res.* **90**, 499–541 (2020).
- Lateef, F. Simulation-based learning: just like the real thing. *J. Emerg. Trauma Shock* **3**, 348–352 (2010).
- Lebert, A. & Vilarroya, Ó. The links between experiential learning and 4E cognition. *Ann. N. Y. Acad. Sci.* **1541**, 37–52 (2024).
- Lee, C. A., Pais, K., Kelling, S. & Anderson, O. S. A scoping review to understand simulation used in interprofessional education. *J. Interprof. Educ. Pract.* **13**, 15–23 (2018).
- Kaplonyi, J. et al. Understanding the impact of simulated patients on health care learners' communication skills: a systematic review. *Med. Educ.* **51**, 1209–1219 (2017).
- Alharbi, A., Nurfianti, A., Mullen, R. F., McClure, J. D. & Miller, W. H. The effectiveness of simulation-based learning (SBL) on students' knowledge and skills in nursing programs: a systematic review. *BMC Med. Educ.* **24**, 1099 (2024).
- Clark, A. D. et al. The effect of 3-dimensional simulation on neurosurgical skill acquisition and surgical performance: a review of the literature. *J. Surg. Educ.* **74**, 828–836 (2017).
- von Wendt, C. E. A. & Niemi-Murola, L. Simulation in interprofessional clinical education: exploring validated nontechnical skills measurement tools. *Simul. Healthc.* **13**, 131 (2018).
- McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H. & Wayne, D. B. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad. Med.* **86**, 706 (2011).
- Hofmann, R., Curran, S. & Dickens, S. Models and measures of learning outcomes for non-technical skills in simulation-based medical education: Findings from an integrated scoping review of research and content analysis of curricular learning objectives. *Stud. Educ. Eval.* **71**, 101093 (2021).
- Manzi, A. & Nigro, G. Long-term memory for performed and observed actions: Retrieval awareness and source monitoring. *Memory* **16**, 595–603 (2008).
- Schult, J., Stölpmagel, R. von & Steffens, M. C. Enactment versus observation: item-specific and relational processing in goal-directed action sequences (and lists of single actions). *PLoS ONE* **9**, e99985 (2014).
- Convertino, G. et al. Positive and negative touch differentially modulate metacognitive memory judgements for emotional stimuli. *Br. J. Psychol.* **116**, 34–51 (2025).
- Della Longa, L., Gliga, T. & Farroni, T. Tune to touch: affective touch enhances learning of face identity in 4-month-old infants. *Dev. Cogn. Neurosci.* **35**, 42–46 (2019).
- Cook, S. W., Mitchell, Z. & Goldin-Meadow, S. Gesturing makes learning last. *Cognition* **106**, 1047–1058 (2008).
- Dijkstra, K., Kaschak, M. P. & Zwaan, R. A. Body posture facilitates retrieval of autobiographical memories. *Cognition* **102**, 139–149 (2007).
- Roberts, B. R. T., MacLeod, C. M. & Fernandes, M. A. The enactment effect: a systematic review and meta-analysis of behavioral, neuroimaging, and patient studies. *Psychol. Bull.* **148**, 397–434 (2022).
- Iani, F. Embodied memories: reviewing the role of the body in memory processes. *Psychon. Bull. Rev.* **26**, 1747–1766 (2019).
- Schnotz, W. An integrated model of text and picture comprehension. in *The Cambridge Handbook of Multimedia Learning* (ed. Mayer, R. E.) Vol. 49–69 (Cambridge University Press, 2005).
- Eitel, A. & Scheiter, K. Picture or text first? Explaining sequence effects when learning with pictures and text. *Educ. Psychol. Rev.* **27**, 153–180 (2015).
- Vogt, A., Babel, F., Hock, P., Baumann, M. & Seufert, T. Immersive virtual reality or auditory text first? Effects of adequate sequencing and prompting on learning outcome. *Br. J. Educ. Technol.* **52**, 2058–2076 (2021).
- Humphreys, M. S., Hockley, W. E. & Chalmers, K. A. Recognition memory: the probe, the returned signal, and the decision. *Psychon. Bull. Rev.* **31**, 568–598 (2024).
- Hurlstone, M. J. Serial recall. In *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications* (eds Kahana, M. J. & Wagner, A. D.) <https://doi.org/10.1093/oxfordhb/9780190917982.013.28> (Oxford University Press, 2024).
- Logan, G. D. & Cox, G. E. Serial order depends on item-dependent and item-independent contexts. *Psychol. Rev.* **130**, 1672–1687 (2023).
- Mack, M. L. & Palmeri, T. J. Discrimination, recognition, and classification. In *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications* (eds Kahana, M. J. & Wagner, A. D.) <https://doi.org/10.1093/oxfordhb/9780190917982.013.30> (Oxford University Press, 2024).
- Tartaglia, E. M., Bamert, L., Herzog, M. H. & Mast, F. W. Perceptual learning of motion discrimination by mental imagery. *J. Vis.* **12**, 14 (2012).
- Marre, Q., Huet, N. & Labeye, E. Embodied mental imagery improves memory. *Q. J. Exp. Psychol.* **74**, 1396–1405 (2021).
- Brodt, S. et al. Rapid and independent memory formation in the parietal cortex. *Proc. Natl. Acad. Sci. USA* **113**, 13251–13256 (2016).
- Kumar, N., Manning, T. F. & Ostry, D. J. Somatosensory cortex participates in the consolidation of human motor memory. *PLoS Biol.* **17**, e3000469 (2019).
- Kruger, J. S., Kruger, D. J. & Suzuki, R. Assessing the effectiveness of experiential learning in a student-run free clinic. *Pedagog. Health Promot.* **1**, 91–94 (2015).
- Waddell, A., Lennox, A., Spassova, G. & Bragge, P. Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. *Implement. Sci.* **16**, 74 (2021).
- Eppich, W. & Cheng, A. Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simul. Healthc.* **10**, 106 (2015).
- Rudolph, J. W., Raemer, D. B. & Simon, R. Establishing a safe container for learning in simulation: the role of the presimulation briefing. *Simul. Healthc.* **9**, 339 (2014).
- Fraser, K. L., Ayres, P. & Sweller, J. Cognitive load theory for the design of medical simulations. *Simul. Healthc.* **10**, 295–307 (2015).
- Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M. & Chen, A.-S. Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *J. Educ. Comput. Res.* **58**, 596–615 (2020).
- Saylik, R., Castiello, S. & Murphy, R. A. The role of emotional interference on learning in an emotional probabilistic Go/No-Go task. *Dusunen Adam J. Psychiatry Neurol. Sci.* **34**, 1 (2021).

39. Colquitt, J. A., LePine, J. A. & Noe, R. A. Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *J. Appl. Psychol.* **85**, 678–707 (2000).
40. Judge, T. A. & Bono, J. E. Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: a meta-analysis. *J. Appl. Psychol.* **86**, 80–92 (2001).
41. Richardson, M., Abraham, C. & Bond, R. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* **138**, 353–387 (2012).
42. Koriati, A. The self-consistency model of subjective confidence. *Psychol. Rev.* **119**, 80–113 (2012).
43. Dunning, D. Chapter five—The Dunning–Kruger Effect: on being ignorant of one's own ignorance. in *Advances in Experimental Social Psychology* Vol. 44 (eds Olson, J. M. & Zanna, M. P.) 247–296 (Academic Press, 2011).
44. Pazin Filho, A. & Romano, M. M. D. Simulação: aspectos conceituais. *Medicina* **40**, 167–170 (2007).
45. Fenwick, T. J. Expanding conceptions of experiential learning: a review of the five contemporary perspectives on cognition. *Adult Educ. Q.* **50**, 243–272 (2000).
46. Itin, C. M. Reasserting the philosophy of experiential education as a vehicle for change in the 21st century. *J. Exp. Educ.* **22**, 91–98 (1999).
47. Kolb, D. A. *Experiential Learning: Experience as the Source of Learning and Development* (FT Press, 2014).
48. Sweller, J. Cognitive load during problem solving: effects on learning. *Cogn. Sci.* **12**, 257–285 (1988).
49. Sweller, J. Cognitive load theory. in *Psychology of Learning and Motivation* Vol. 55, 37–76 (Elsevier, 2011).
50. Sweller, J., van Merriënboer, J. J. G. & Paas, F. Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* **31**, 261–292 (2019).
51. Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res.* **44**, 314–324 (2012).
52. Hecht, H., Welsch, R., Viehoff, J. & Longo, M. R. The shape of personal space. *Acta Psychol.* **193**, 113–122 (2019).
53. Lebert, A., Vergilino-Perez, D. & Chaby, L. Keeping distance or getting closer: How others' emotions shape approach-avoidance postural behaviors and preferred interpersonal distance. *PLoS ONE* **19**, e0298069 (2024).
54. Grèzes, J., Risch, N., Courtet, P., Olié, E. & Mennella, R. Depression and approach-avoidance decisions to emotional displays: the role of anhedonia. *Behav. Res. Ther.* **164**, 104306 (2023).
55. Knyazev, G. G., Bocharov, A. V., Slobodskaya, H. R. & Ryabichenko, T. I. Personality-linked biases in perception of emotional facial expressions. *Personal. Individ. Differ.* **44**, 1093–1104 (2008).
56. Butler, E. A. et al. The social consequences of expressive suppression. *Emotion* **3**, 48–67 (2003).
57. Topp, C. W., Østergaard, S. D., Søndergaard, S. & Bech, P. The WHO-5 Well-Being Index: a systematic review of the literature. *Psychother. Psychosom.* **84**, 167–176 (2015).
58. Buéla-Casa, G., Guillén-Riquelme, A. & Seisdedos Cubero, N. Cuestionario de Ansiedad Estado-Rasgo: Adaptación española. (TEA Ediciones, 2011).
59. Espejo, B., Checa, I., Perales-Puchalt, J. & Lisón, J. F. Validation and measurement invariance of the Scale of Positive and Negative Experience (SPANE) in a Spanish general sample. *Int. J. Environ. Res. Public Health* **17**, 8359 (2020).
60. Rempel, D., Willms, K., Anshel, J., Jaschinski, W. & Sheedy, J. The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Hum. Factors* **49**, 830–838 (2007).
61. RStudio Team. RStudio: integrated development for R. RStudio, PBC. (2020).
62. Tukey, J. W. *Exploratory Data Analysis* (Addison-Wesley, 1977).
63. Zellner, A., Richardson, A. M., Lidbury, B. A., Hobson, P. & Badrick, T. An investigation into outlier elimination and calculation methods in the determination of reference intervals using serum immunoglobulin A as a model data collection. Preprint at <https://doi.org/10.48550/arXiv.1907.09637> (2019).
64. Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49**, 1373–1379 (1996).
65. Van Smeden, M. et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat. Methods Med. Res.* **28**, 2455–2474 (2019).

Acknowledgements

The authors thank Laura Alsina, Mila Markovic, and Anna Yue for their assistance during the experimental sessions, as well as the actors Mar Casas, Berta Graells, Mirena Nafarrete, Arnau Armengol, and Carla Vilaró for their dedication. Special thanks to Dr. Meritxell Girvent and Dr. Carolina Pérez García for their assistance in designing the simulation protocol, and to Dr. Daniel Bergé for his contribution to developing the fictitious clinical syndromes. This work was supported by Fundación Tatiana. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

A.L.: conceptualization, methodology, software, formal analysis, investigation, writing. O.V.: conceptualization, methodology, supervision, project administration, resources, funding acquisition, writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41539-025-00380-9>.

Correspondence and requests for materials should be addressed to Óscar Vilarroya.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025