Erasmus+

SCALENEo

Score Your Way to Clinical Reasoning Excellence: SCALENEo Online Serious Game in Physiotherapy Education

Co-funded by the European Union

Cofinancé par l'Union européenne

Renaud Hage[1,2,3], Frédéric Dierick[1,3,4], Joël Da Natividade[4], Simon Daniau[5], Wesley Estievenart[1], Sébastien Leteneur[6], Jean-Christophe Servotte[5], Mark Jones[7,8], Fabien Buisseret[1]

[1]Centre de Recherche et de Formation continue de la HELHa (CeREF), Chaussée de Binche 159, 7000 Mons, Belgium ;

[2] Traitement Formation Thérapie Manuelle (TFTM), Manual Therapy Center, Avenue des Cerisiers 211A, 1200 Brussels, Belgium ;

[3] Faculté des Sciences de La Motricité, UCLouvain, Place Pierre de Coubertin 1-2, 1348 Louvain-la-Neuve, Belgium ;

[4] RehaLAB, Centre National de Rééducation Fonctionnelle et de Réadaptation—Rehazenter, Rue André Vésale 1, 2674 Luxembourg, Luxembourg ;

[5] Fors Recherche & Développement, Haute école de Namur-Liège-Luxembourg (Henallux) ;

[6] Université Polytechnique Hauts-de-France, LAMIH, CNRS, UMR 8201, F-59313 Valenciennes, France.

[7]Allied Health and Human Performance Academic Unit, University of South Australia, Adelaide, Australia;

[8]International Centre for Allied Health Evidence, University of South Australia, Adelaide, Australia.

Abstract

Background: SCALENEo (Smart ClinicAL rEasoning iN physiothErapy) is an innovative online serious game designed to improve clinical reasoning in musculoskeletal physiotherapy education. Adapted from the "Happy Families" card game, it provides an interactive, structured approach to developing learners' ability to categorize clinical information into families of hypotheses.

Methods: SCALENEo supports both self-directed and collaborative learning, eliminating the need for continuous instructor supervision while ensuring meaningful engagement.

Results: SCALENEo's feedback and scoring system assesses learners' decision-making processes but also promotes structured and reflective evidence-based reasoning over random guessing

Conclusion: SCALENEo represents an effective tool for clinical reasoning education in musculoskeletal physiotherapy. Future research will explore its long-term impact on clinical decision-making skills.

Keywords: *Clinical reasoning, serious games, physiotherapy education, gamification, digital learning*

Background

Learning clinical reasoning (CR) is an integral component of physiotherapy education (Schuwirth, 2009) and professional practice, directly influencing decision-making processes crucial for effective and safe patient care. CR encompasses a range of analytical and non-analytical processes and skills used in patient assessment and management, involving both the development of hypotheses and pattern recognition (Croskerry, 2009a, 2009b; Jones, 1992). For physiotherapy students, mastering hypothetico-deductive analytical reasoning - involving generating and testing multiple hypotheses - is essential due to the variability and complexity of clinical presentations (Bowen, 2006; Jones, 1992; Jones, 2019; Rivett, 2004).

The importance of employing different categories of hypotheses (Jones, 2019), also known as families of hypotheses (Hage, 2023), is well-documented in the reflective practices of experienced clinicians and educators (Barlow, 2012; Edwards, 2004; Jones, 2019; Rushton, 2016; Smart & Doody, 2007). This structured approach enhances clinical judgment consistency, fosters thorough analysis of patients' cases, and mitigates cognitive biases. However, despite its recognized value, teaching and assessing CR remains challenging. Traditional methods, such as case-based discussions or direct supervision, are time-consuming and may not always provide interactive and adaptive learning opportunities suited to diverse student needs.

In response to these challenges, a variety of tools and methods are needed in educational settings to optimize learning for the diverse profiles and skills students possess. One promising approach is gamification, which can enhance learning engagement and cooperation by applying real-world scenarios (Krishnamurthy, 2022). Specifically, serious games (SG), a type of game-based learning aimed at achieving specific educational outcomes (Gorbanev, 2018; Haoran, 2019), have shown efficacy in improving concentration and knowledge retention among learners (García-Redondo, 2019; Gorbanev, 2018; Tubelo, 2019). CR learning through SG (Blanié, 2020; Koivisto, 2018) and online simulation games (Cant, 2010; Lapkin, 2010) has already been successfully implemented in nursing education, showing promising results in improving hypothesis-driven analysis and structured decision-making.

The integration of SG in nursing (Koivisto, 2018) and medical education (Tubelo, 2019) has yielded positive outcomes (Qiao et al., 2023). Inspired by these successes, we extended the use of SG learning to physiotherapy education through the "Happy Families" card game (Hage, 2022; 2023). This card game appears to be a promising tool for engaging students in learning CR and improving their awareness of cognitive biases and reflective practice. The "Happy Families" SG focuses mainly on cognitive bias processes through the perception of relevant or significant information when analyzing a clinical case using hypothetico-deductive reasoning. Moreover, because the social and cultural environment in learning is important, it also takes place in a social framework that is supported by socio-constructivist learning theory (Clark & Dumas, 2015; Hickey, 1997). In fact, learning and the motivation to learn are reinforced by exchanges within groups of learners. At the start of the game, players (i.e. students, typically in groups of 3 to 5 players with a maximum of 5 groups per session) are provided with a fully detailed clinical case including both quantitative and qualitative patient data. The game includes a set of cards representing the hypotheses families: core categories of clinical judgment such as patient's expectations, pain, origin of symptoms etc, see Hage et al. (2022) for a full explanation. Students must then select the cards where they have initial hypotheses and, using an erasable pen, write a list of their hypotheses on the back of the card. In addition, a "Wild card" is provided for players to write their main hypotheses (i.e. first choice at this stage) in each family. The game advances as players categorize hypotheses and refine their reasoning. The game concludes once all hypotheses have been categorized. Then each group presents its analysis to the other groups, allowing them to discuss, justify and compare their hypotheses. Interaction between the players and feedback from the teacher help to refine or modify the hypotheses.

Beyond SG, technology-enhanced learning is increasingly recognized as a vital tool for improving both education and continuing professional development, which are essential for ensuring high-quality healthcare services. The positive impact of using virtual scenarios as learning activities on outcomes such as knowledge acquisition and CR has been well-established in medicine (Kononowicz et al., 2019). These technological methods can be described as "interactive computer simulations based on clinical scenario cases". In a 2019 systematic review and meta-analysis assessing the effectiveness of virtual patients compared to traditional teaching, Kononowicz et al. (2019) found that using virtual patients improved or equaled traditional education in terms of CR, procedural skills, and a combination of procedural and team skills.

The Erasmus+ "SCALENEo" (Smart ClinicAL rEasoning iN physiothErapy) project (https://scaleneo.eu) aims to promote the SG "Happy Families" through an online approach. The aim is to leverage the possibilities offered by information and communication technologies to develop an online version of this SG to reach geographically remote and emerging audiences and, mostly, to allow individual players, i.e. students or physiotherapists participating in ongoing professional development of CR. While maintaining the same theoretical concepts, online gaming works a little differently to the face-to-face game. In the face-to-face game form, participants have to list the information related to each hypothesis family, so the same information can be found in different families of hypotheses (Fig1A). In the online version, the sentences explaining the clinical case can be linked to the hypothesis families (Fig1B). The online implementation of the SG leads to a key question which is at the heart of the present study: How to provide relevant feedback about one player's CR? As in test-enhanced learning (Roediger, 2006), it is indeed crucial to evaluate students and to provide feedback on the quality of their work and their progress in learning CR. Thanks to evaluation, the Happy Families game can therefore be used as a formative evaluation tool, provided that the students can be evaluated.

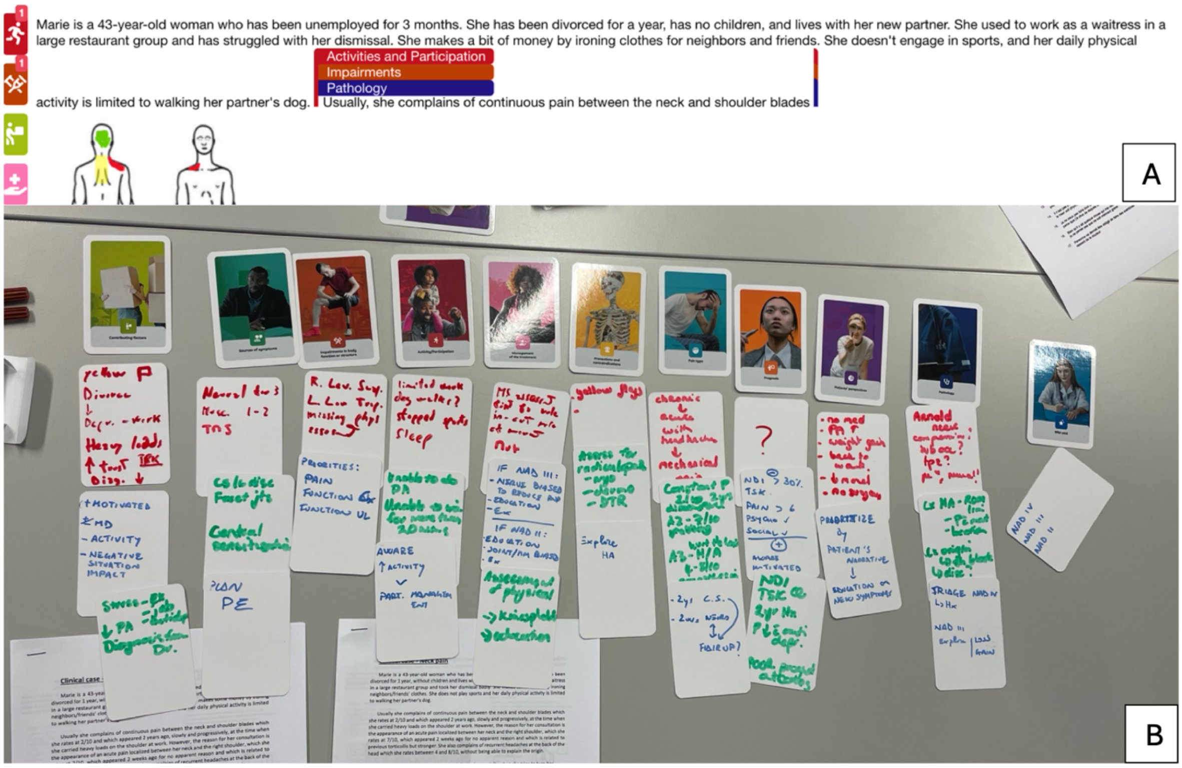


Figure 1. A: online version, the sentences related to the clinical case are assigned by the participant to one or more hypothesis families. B: the "face to face" version showing all families with information (e.g. hypothesis) from the clinical case written by participants. Note: The pictures are only intended to display the SG under different forms, the answers given are not necessarily correct.

The value of student assessment is well established (Larsen, 2008). CR skills are particularly challenging to assess because they encompass a combination of cognitive abilities required to analyze and synthesize problems (Min Simpkins, 2019). Research in cognitive psychology indicates that the use of tests can enhance understanding and application (Larsen, 2008). Evaluating CR, which stems from complex internal cognitive processes that are not directly observable, is inherently problematic. CR involves the integration of cognitive, psychomotor, and affective skills (Huhn, 2019), necessitating a flexible assessment procedure. The term CR encompasses various skills including decision-making, critical thinking, problem-solving, clinical judgment, and diagnostic reasoning (Young, 2019). Therefore, it is through these terms that we must identify and assess the relevant skills that will eventually lead to a global assessment and score.

Despite the demonstrated benefits of SG in nursing and medicine fields, there is a notable lack of research on the use of such tools for assessing and enhancing CR in physiotherapy. More, in the medical field, there is no gold standard for assessing CR (Simpkins, 2019) and this is also true in physiotherapy, highlighting a significant gap in the literature. Current assessment methods in medical education, such as the Assessment of Reasoning Tool (Thammasitboon, 2018), the Think Aloud protocols (Pinnock, 2015) and the Script Concordance Test (SCT) (Charlin, 2000) have been proposed. Specifically, the SCT method has been used since the work of Charlin et al. (2000). The SCT is based on script theory and involves comparing students' interpretation of clinical data with those of experienced clinicians through a series of clinical tasks presented in specific and "ill-defined" contexts. In summary, an ill-defined case is presented to candidates as a brief clinical scenario where the information provided is insufficient to make a definitive decision. In addition, data provided by motion capture systems are integrated to the scenario to improve their integration in the participants' initial CR processes. Thus, biomechanical data (e.g. kinematic values) that illustrate the clinical case are presented in connection with the clinical case. The objective is to enrich the participant's reasoning with quantitative, sensor-based, data. Each case is followed by several questions that begin with a hypothesis, followed by additional information. Participants are then asked to assess how this new information impacts the likelihood of the proposed hypothesis being correct (Brentnall, 2022; Charlin, 2000; Dory, 2012, 2016).

Since the SCT method has proved reliable and effective for measuring CR in health professions such as nursing (Deschênes, 2011), medicine (Cohen Aubart, 2021) and physiotherapy (Kojich, 2024; O'Brien, 2023), the aim of this study was therefore to show if an adaptation of the basic principle of SCT is feasible for testing physiotherapy CR during the "Happy Families" game. As in the original SCT, the information added progressively by the person who prepared the clinical case will have to be confirmed or disproved. Players must determine whether additional information reinforces an existing hypothesis or necessitates the creation of a new one. he players are then asked to assess the impact of the information on the probability that the hypotheses proposed are correct. By integrating the adaptation of the SCT in an online version of the “Happy Families” card game, we aim to enhance the quality of CR training and ultimately improve patient physiotherapy care outcomes.

Methods

First, a description of how to play the game will be given (https://scaleneo.eu/en/home-english/). Then the technical details needed to understand the feedback will be provided.

Online game

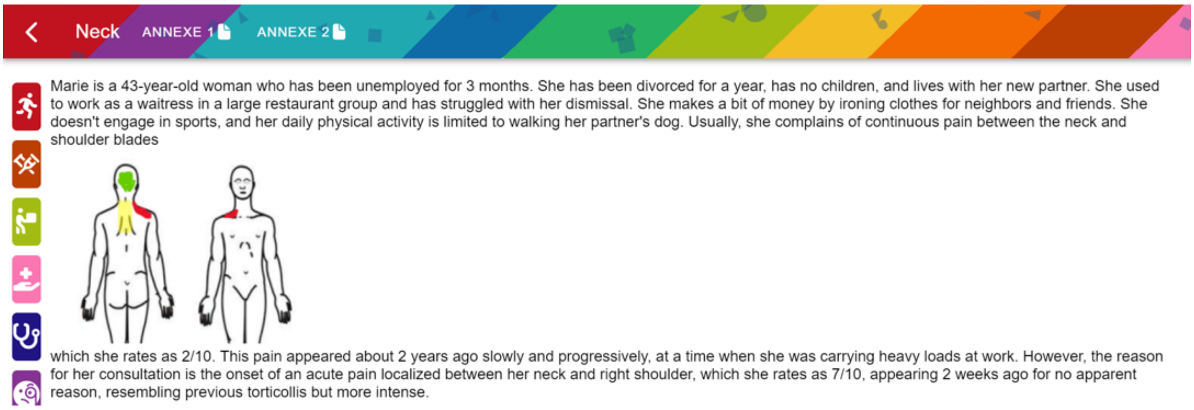
The game starts when a player or a group of player selects a given clinical case, the clinical cases being labelled by pain localization (e.g. neck pain, hip pain, etc). The player(s) then read the clinical case, mostly composed of text and biomechanical data. Pre-selected groups of words or elements can be, by clicking, selected and assigned to one or more hypothesis families in the form of a card (Fig. 2A). The player can select an element of information (an item) by clicking once to mark the item as relevant or clicking twice to mark it as significant, then assigning it to a hypothesis family card. Significant means essential elements in the reference game that players must identify and highlight as critical. This first phase of the game therefore requires hypothetico-deductive reflection on the clinical case and the selection and linking of elements of information to the hypothesis family or families the player(s) judge relevant. It has to be said that, for each clinical cases, a reference game exists, i.e. a “perfect” game obtained by consensus by the authors of the present paper. In this reference game, the authors have not only classified relevant information into the appropriate hypothesis families but also identified the items of information by importance.

During the game, a chat is enabled to allow for discussion in the case of a multiplayer game. Players are allowed to consult their courses or any relevant source of information – articles published in peer-reviewed journals are also suggested in the clinical cases.

When the player(s) has/have ended their analysis, all selections and assignments are made and a blue button on the right corner of the screen can be clicked that triggers the display of the results (Fig. 3A). The first feedback (Fig. 3B) is a hypothesis-by-hypothesis count of the information correctly assigned by the player(s). A green/orange/red light is related to the percentage of items correctly assigned and the correct identification of significant information, see below for a detailed explanation. Figure 3C finally shows a global evaluation using a gold/silver/bronze brain symbol and the corresponding total score. Three difficulty levels are implemented, giving the player(s) the opportunity to make several attempts before receiving definitive feedback depending on the level of difficulty: easy is 3 attempts, medium is 2 and hard is 1. The global brain level score is calculated according to the number of attempts needed to eliminate all the red lights: first attempt, gold, second attempt, silver, third attempt, bronze. If, after the allowed attempts, there are still cards with red lights, no brain level is awarded, which signifies the player(s)'s failure to identify and classify all the significant elements.

At the end of all attempts, a wild card must be selected (Fig. 4A). This card represents, according to the player(s), the key information guiding the clinical reasoning and physical management of the presented case. All hypothesis categories must be considered, and those related to precautions and contraindications may be prioritized when necessary. A final score is then computed. After the choice of the wild card, the final feedback is displayed (Fig. 4B): It consists of an overall score and of all the other elements mentioned above. At this stage it is also possible to look at the reference game. The total score is based on the average card score for the families of clinical judgment considered most important by the authors of this study. For example, if the hypotheses "type of pain", "sources of symptoms" and "precautions and contraindications" were judged to be most important to the selected clinical case, then the total score would be based on those families. The total score is updated after the last attempt and is now the average score between the 3 above families and the wild card score.

A



B

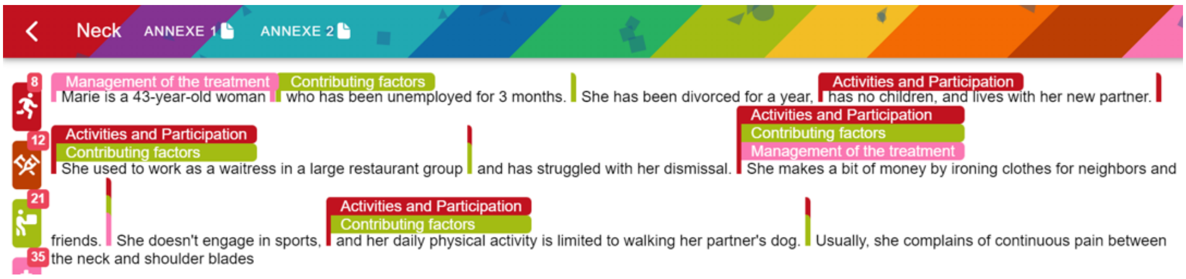


Figure 2. Screenshot of a neck-pain clinical case. The clinical case is displayed before assignments of the information to hypothesis families (A) and after (B). Note: The assignments in panel B are not necessarily correct and are taken from a random group of player's game.

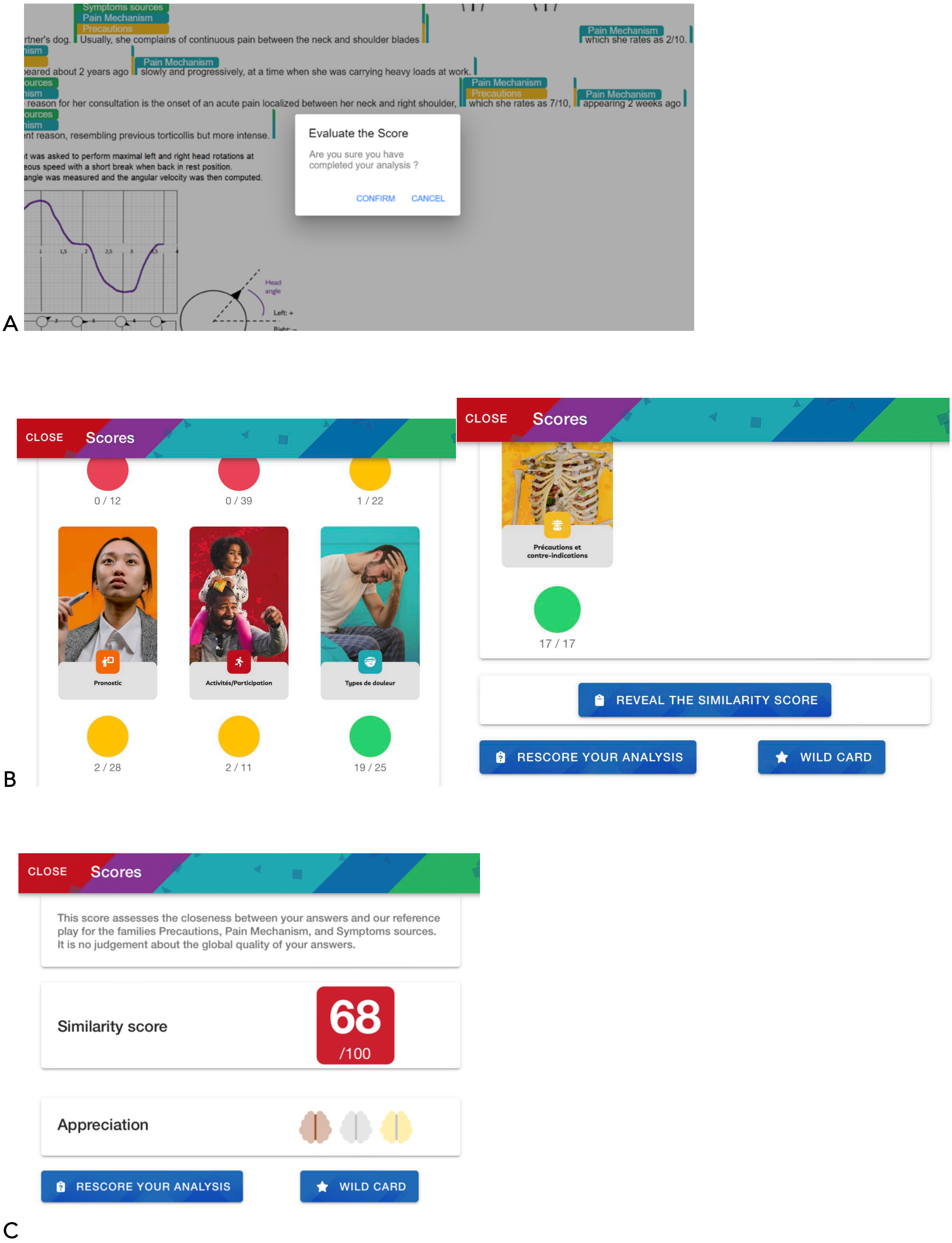


Figure 3. A: End of a game and beginning of the assessment procedure. B: Hypothesis-by-hypothesis assessment; counting of the number of elements correctly assigned by the player(s). C: The score is revealed with a warning message and a global appreciation under the form of a gold/silver/bronze brain.

Figure 4. A: Selection of the wild card. B: Final feedback. The score here is very low since the image has been generated from a randomly generated game.

Classifying card elements

We now turn to the technical issues "hidden" behind the online game; these points are a key part of the novel approach we propose to assess CR.

Technically, a played game consists of a set of selected items by the player(s) and the game cards that must be compared to the reference cards, i.e. the cards filled by the SCALENEo experts.

Table 1: Schematic view of one element's classification according to its presence or not in a game card. Relevant items are those present in the reference game but not considered significant. Significant items, on the other hand, are essential elements in the reference game that players must identify and highlight as critical. Good items are those that have been correctly assigned by the player(s) to the appropriate hypothesis family. Missed items refer to those present in the reference game but not identified or selected by the player(s). Wrong items are those selected by the player(s) but not present in the reference game. Bad items are those selected by the player(s) but incorrectly assigned to the wrong category (significant instead of relevant or relevant instead of significant). Finally, empty items are those that are neither selected by the player(s) nor present in the reference game.

		Relevant/Significant Item in Reference card	
		Present	Not present
Relevant/Significant Item in Game card	Present	Good/bad	wrong/
	Not present	missed	empty

Lights and thresholds

After comparison between reference and game cards, the lights are displayed. Their color is:

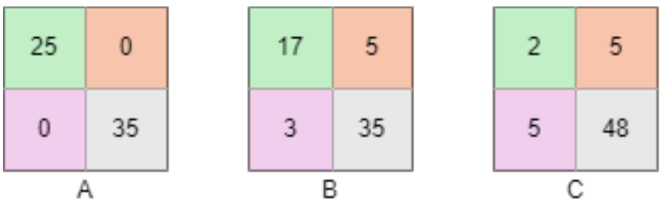
- Green if the game and reference cards are either both empty or if all the significant elements are found and more than 50% of the reference card elements are correctly identified. This threshold of 50% is a parameter and could be lowered or increased according to the player(s) experience for example.

- Yellow if all the significant elements are found and the percentage of correctly identified items is below the threshold of 50%.
- Red if significant elements have been missed or classified in the wrong family card or incorrectly marked as relevant.

Total score

The total score is based on the average card score for the families of judgment considered most important by the authors of this study. To compute the card score, a Cohen's Kappa coefficient measuring the agreement between the played game and the reference one is computed - the game and reference cards are therefore seen as two independent observers whose agreement can be measured. Cohen's Kappa is finally multiplied by 100 to have a card score between 0 and 100. Note that if a card in the reference game is empty and the corresponding card in the game played is also empty, the score is defined to be 100%, because the two observers agree not to put any elements in the cards. Figure 4 displays 3 examples of card score calculations.

Figure 4. Examples of distribution of elements in a card according to the classification described in Table 1. A: Only good items present in both cards, leads to a score of 100%. B: Some missed or wrong items are present; this example leads to score of 71%. C: This example leads to a score of 19 %, because of low number of good items.



Score evolution vs number of errors

The following procedure is proposed to assess our way of scoring a game. First, using a developer tool in the online application, game samples were generated by duplicating one clinical case's reference game (neck pain patient), and then by introducing errors by randomly removing a given percentage (*n)* of items. Twenty samples for each value of *n* were generated with *n* = 0, 5, 10, …, 100 %. Second, the same procedure was applied but items were randomly replaced by a random item instead of being removed. A total of 840 plays has been generated. In both cases, at a given *n*, interpreted as the % of errors, the average score and its standard deviation were computed for a single attempt (no wild card choice has been performed). Then, the score decrease rate was computed as the derivative of the score as a function of *n*, where the latter function of *n* was obtained by a fourth-order regression line.

The games with removed items are interpreted as representative of a "cautious player", i.e. a player preferring not to answer when he/she is not certain of the answer. The games with replaced items are interpreted as representative of a "bold player", i.e. a player taking the risk of answering to items of which they are not fully sure.

Results

The evolution of the average score versus the percentage of error is displayed in Fig. 6, where it can be observed that the scores with replaced items are always lower than with removed items at a given percentage of error: The bold player has always a lower score than the cautious one in this approach.

The score decrease rate versus the percentage of error is displayed in Fig. 7. The score is not very sensitive to the % of error if the latter is small for a cautious player. Both cautious and bold payers' scores are strongly impacted by the % of error as soon as the latter is larger than about 25 %. The numerical values are given in Appendix, see Tables A1 and A2 for completeness.

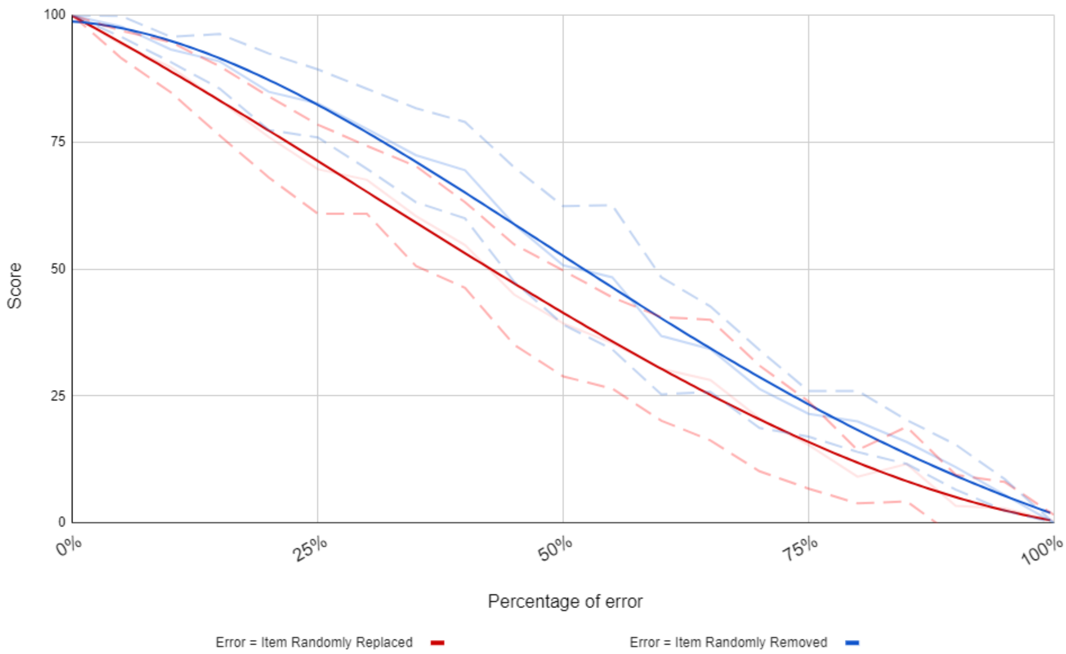


Figure 6. Evolution of the averaged scores (computed on 20 plays) versus the percentage of error. The blue line shows the generated plays with removed items, and the red line shows the generated plays with replaced items. The dashed lines define an interval of ± 1 standard deviation around the average scores. The blue and red lines show a fourth-order regression line of the averaged scores to better guide the eyes. For completeness we mention that they have equations Score = $98.8 - 14.1\ x - 266\ x^2 + 254\ x^3 - 71.1\ x^4$ and Score = $99.9 - 104\ x - 58.1\ x^2 + 65.3\ x^3 - 2.8\ x^4$ respectively, with x = Percentage of error.

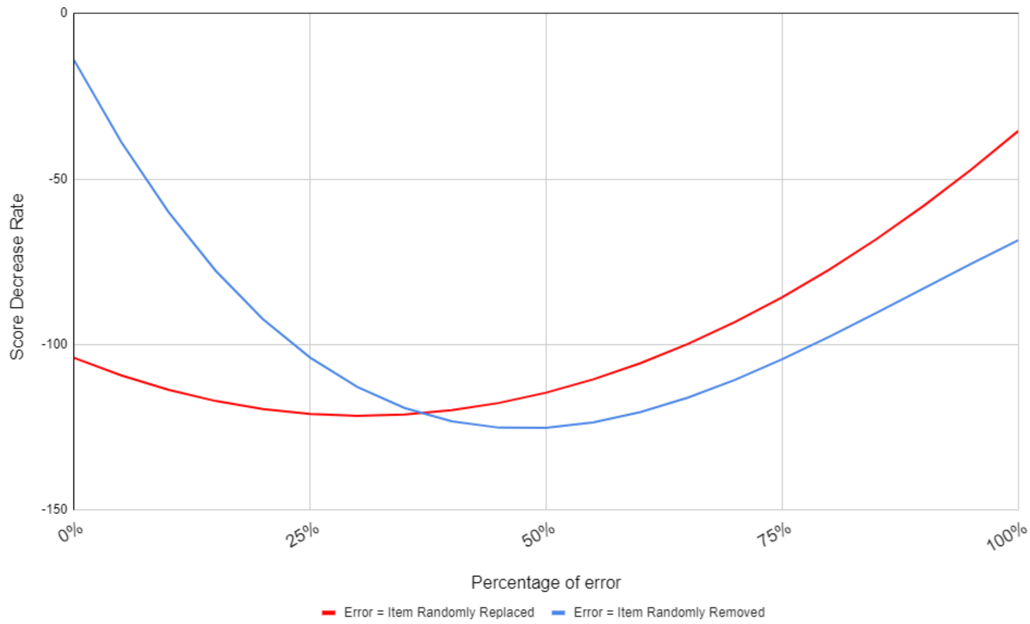


Figure 7. Same as Fig. 6 for the score decrease rate, computed as the derivative of the score regression lines versus n. The negative Score Decrease Rate observed at a given percentage of error can be seen as the score variation corresponding to a unitary increase of percentage of error.

Discussion

Our SG shares similarities with SCT, a well-established method for assessing CR. In SCT, participants assess a clinical situation using limited information, after which additional information is introduced to reassess their previous choices. These are then evaluated against expert judgments (Dory et al. 2016, Charlin et al. 2000). Conversely, in SCALENEo, players must identify and categorize information from a comprehensive clinical case presented upfront. This design fosters active engagement, encourages players to justify their thought processes, and strengthens their ability to navigate complex and ill-defined scenarios - a key aspect of effective CR (Croskerry, 2009b). In our clinical cases, each card can be considered as a "case" that will need to be categorized according to each family of hypotheses. That is why the first feedback we provide is the collection of lights, displayed card-by-card. This aims at showing the player(s) which hypotheses families they understand satisfactorily (or poorly or not) and, maybe most important is that the clinical red flags are clearly identified so students with this SG prepare for first line in clinical practice. Red flags are used to help clinicians identify serious pathology, and the majority of guidelines recommend the use of red flags (Finucane et al., 2020). One of SCALENEo's key innovations is its two-steps feedback system: (1) card-by-card feedback; and (2) global scoring.

The integration of card-by-card feedback aligns with prior research emphasizing the value of formative assessment in CR education (Larsen, 2008). Immediate, targeted feedback allows learners to identify cognitive gaps and improve iteratively, promoting deeper understanding and reflective practice. This approach mirrors the structured feedback mechanisms in SCT but innovates by embedding it in an engaging gamified format. Gamification, as highlighted by Koivisto et al. (2018), provides principles for designing educational tools that enhance engagement and learning outcomes. SCALENEo leverages these principles effectively, fostering motivation and active participation among learners.

Our results demonstrate the value of scoring as a secondary, rather than primary, feedback mechanism. While scoring is not central in face-to-face sessions led by expert teachers, it is essential for unsupervised play in SCALENEo. The global scoring reinforces cautious decision-making, rewarding those who analyze cases methodically rather than guessing. This approach aligns with educational theories that emphasize the importance of careful hypothesis testing (Larsen et al., 2008) and the mitigation of cognitive biases in clinical judgment (Croskerry, 2009b). It is particularly suited for intermediate learners who possess foundational knowledge of hypothesis families and are ready to refine their skills further. Because physiotherapy requires a broad base of skills and knowledge to enable the correct identification of clinical priorities and the application of appropriate medical interventions, we are satisfied with our scoring system that appears to favor caution: the risk-taker player always scores lower than the cautious player in this approach. This may teach students to always favor caution when assessing a clinical case, hoping that they can replicate this with real patients. The study of score evolution versus error rates underscores the importance of preliminary instruction in hypothesis families. The quick increase in scores for players with error rates below 50% suggests that SCALENEo is most effective when integrated into a broader curriculum with pre-game theoretical modules. This structured preparation ensures learners can engage meaningfully with the game's content, supporting both knowledge construction and skill application (Gorbanev et al., 2018). This structuring should increase players' motivation, particularly by encouraging them to replay games. This can be explained by two reasons: 1) the feeling of competence and 2) the feeling of control over the task to be performed (Viau, 2009). I have placed it as a comment in the document. Honestly, check if you want to add it or not.

Despite its strengths, certain limitations of our scoring method of the SCALENEo game must be acknowledged. The reliance on expert-defined reference games assumes homogeneity in CR, potentially overlooking variations in reasoning styles or regional practice differences. Future developments could explore adaptive feedback and dynamic difficulty adjustments to make the game more inclusive for novice learners while maintaining its value for intermediate and advanced users. Incorporating design principles from Koivisto et al. (2018) could further refine the game's accessibility and adaptability: as mentioned above, depending on the clinical case, different families of hypotheses could be deemed most important, and the score can therefore be adaptive. Additionally, the online platform offers significant opportunities for such research. Anonymized gameplay data could enable large-scale analyses of CR patterns across different demographics and educational backgrounds. Such insights could guide targeted instructional material development and address gaps in existing curricula. For example, differences in reasoning approaches between novices and experts, as outlined by Croskerry (2009b), could inform personalized feedback mechanisms in SCALENEo.

Lastly, in terms of perspectives, SCALENEo's gamified approach has broader implications beyond musculoskeletal physiotherapy. By adapting the game to other domains, such as neurological or pediatric physiotherapy, its benefits could be extended to a wider range of learners. This scalability highlights its potential as a transformative tool in healthcare education. Furthermore, its emphasis on mitigating cognitive biases and encouraging cautious decision-making underscores its alignment with the principles of safe and effective patient care (Croskerry, 2009b).

In conclusion, our scoring method of the SCALENEo game combines the strengths of serious games with the rigor of established CR assessment methods. Its innovative approach to formative feedback, combined with an engaging and accessible online platform, positions it as a valuable tool for improving CR skills in physiotherapy education. By addressing cognitive biases, leveraging gamification principles, and integrating adaptive feedback, SCALENEo bridges the gap between knowledge and application in a way that traditional methods often fail to achieve.

Funding: This research was Funded by the European Union (Erasmus+ SCALENEo, 22PCS0007).

References

Barlow, S. (2012). "Physiotherapy outpatient's chronic pain management .... Realizing the potential". *NSW Government*.

Blanié, A., Amorim, M.-A., & Benhamou, D. (2020). Comparative value of a simulation by gaming and a traditional teaching method to improve clinical reasoning skills necessary to detect patient deterioration : A randomized study in nursing students. *BMC Medical Education*, *20*(1), Article 1. https://doi.org/10.1186/s12909-020-1939-6

Bowen, J. L. (2006). Educational Strategies to Promote Clinical Diagnostic Reasoning. *The New England Journal of Medicine*.

Brentnall, J., Thackray, D., & Judd, B. (2022). Evaluating the Clinical Reasoning of Student Health Professionals in Placement and Simulation Settings : A Systematic Review. *International Journal of Environmental Research and Public Health*, *19*(2), 936. https://doi.org/10.3390/ijerph19020936

Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education : Systematic review. *Journal of Advanced Nursing*, *66*(1), 3-15. https://doi.org/10.1111/j.1365-2648.2009.05240.x

Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & Van Der Vleuten, C. (2000). The Script Concordance Test : A Tool to Assess the Reflective Clinician. *Teaching and Learning in Medicine*, *12*(4), 189-195. https://doi.org/10.1207/S15328015TLM1204_5

Cohen Aubart, F., Papo, T., Hertig, A., Renaud, M.-C., Steichen, O., Amoura, Z., Braun, M., Palombi, O., Duguet, A., & Roux, D. (2021). Are script concordance tests suitable for the assessment of undergraduate students? A multicenter comparative study. *La Revue de Médecine Interne*, *42*(4), 243-250. https://doi.org/10.1016/j.revmed.2020.11.001

Croskerry, P. (2009a). A Universal Model of Diagnostic Reasoning: *Academic Medicine*, *84*(8), 1022-1028. https://doi.org/10.1097/ACM.0b013e3181ace703

Croskerry, P. (2009b). Clinical cognition and diagnostic error : Applications of a dual process model of reasoning. *Advances in Health Sciences Education*, *14*(S1), 27-35. https://doi.org/10.1007/s10459-009-9182-2

Deschênes, M.-F., Charlin, B., Gagnon, R., & Goudreau, J. (2011). Use of a script concordance test to assess development of clinical reasoning in nursing students. *The Journal of Nursing Education*, *50*(7), 381-387. https://doi.org/10.3928/01484834-20110331-03

Dory, V., Gagnon, R., Charlin, B., Vanpee, D., Leconte, S., Duyver, C., Young, M., & Loye, N. (2016). In Brief : Validity of Case Summaries in Written Examinations of Clinical Reasoning. *Teaching and Learning in Medicine*, *28*(4), 375-384. https://doi.org/10.1080/10401334.2016.1190730

Dory, V., Gagnon, R., Vanpee, D., & Charlin, B. (2012). How to construct and implement script concordance tests : Insights from a systematic review. *Medical Education*, *46*(6), 552-563. https://doi.org/10.1111/j.1365-2923.2011.04211.x

Edwards, I., Jones, M., Carr, J., Braunack-Mayer, A., & Jensen, G. M. (2004). Clinical Reasoning Strategies in Physical Therapy. *Physical Therapy*, *84*(4), 312-330. https://doi.org/10.1093/ptj/84.4.312

Finucane, L. M., Downie, A., Mercer, C., Greenhalgh, S. M., Boissonnault, W. G., Pool-Goudzwaard, A. L., Beneciuk, J. M., Leech, R. L., & Selfe, J. (2020). International Framework for Red Flags for Potential Serious Spinal Pathologies. *Journal of Orthopaedic & Sports Physical Therapy*, *50*(7), 350-372. https://doi.org/10.2519/jospt.2020.9971

García-Redondo, García, Areces, Núñez, & Rodríguez. (2019). Serious Games and Their Effect Improving Attention in Students with Learning Disabilities. *International Journal of Environmental*

Research and Public Health, *16*(14), 2480. https://doi.org/10.3390/ijerph16142480

Gorbanev, I., Agudelo-Londoño, S., González, R. A., Cortes, A., Pomares, A., Delgadillo, V., Yepes, F. J., & Muñoz, Ó. (2018). A systematic review of serious games in medical education : Quality of evidence and pedagogical strategy. *Medical Education Online*, *23*(1), 1438718. https://doi.org/10.1080/10872981.2018.1438718

Hage, R., Fourré, A., Ramonfosse, L., Leteneur, S., Jones, M., & Dierick, F. (2023). Description and rules of a new card game to learn clinical reasoning in musculoskeletal physiotherapy. *Journal of Manual & Manipulative Therapy*, *31*(4), 287-296. https://doi.org/10.1080/10669817.2022.2132346

Haoran, G., Bazakidi, E., & Zary, N. (2019). Serious Games in Health Professions Education: Review of Trends and Learning Efficacy. *Yearbook of Medical Informatics*, *28*(01), 240-248. https://doi.org/10.1055/s-0039-1677904

Huhn, K., Gilliland, S. J., Black, L. L., Wainwright, S. F., & Christensen, N. (2019). Clinical Reasoning in Physical Therapy : A Concept Analysis. *Physical Therapy*, *99*(4), 440-456. https://doi.org/10.1093/ptj/pzy148

Jones, M. A. (1992). Clinical Reasoning in Manual Therapy. *Physical Therapy*, *72*(12), 875-884. https://doi.org/10.1093/ptj/72.12.875

Jones, M. A., Rivett, D. A., & Moore, A. (2019). *Clinical reasoning in musculoskeletal practice* (Second edition). Elsevier.

Jones Mark; Rivett Darren. (2004). *Clinical Reasonning for Manual Therapists* (Butterworth Heinemann). Elsevier Science Limited.

Koivisto, J.-M., Haavisto, E., Niemi, H., Haho, P., Nylund, S., & Multisilta, J. (2018). Design principles for simulation games for learning clinical reasoning: A design-based research approach. *Nurse*

Education Today, *60*, 114-120. https://doi.org/10.1016/j.nedt.2017.10.002

Kojich, L., Miller, S. A., Axman, K., Eacret, T., Koontz, J. A., & Smith, C. (2024). Evaluating clinical reasoning in first year DPT students using a script concordance test. *BMC Medical Education*, *24*(1), Article 1. https://doi.org/10.1186/s12909-024-05281-w

Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Tudor Car, L., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual Patient Simulations in Health Professions Education : Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *Journal of Medical Internet Research*, *21*(7), e14676. https://doi.org/10.2196/14676

Krishnamurthy, K., Selvaraj, N., Gupta, P., Cyriac, B., Dhurairaj, P., Abdullah, A., Krishnapillai, A., Lugova, H., Haque, M., Xie, S., & Ang, E.-T. (2022). Benefits of gamification in medical education. *Clinical Anatomy*, *35*(6), 795-807. https://doi.org/10.1002/ca.23916

Lapkin, S., Levett-Jones, T., Bellchambers, H., & Fernandez, R. (2010). Effectiveness of Patient Simulation Manikins in Teaching Clinical Reasoning Skills to Undergraduate Nursing Students : A Systematic Review. *Clinical Simulation in Nursing*, *6*(6), e207-e222. https://doi.org/10.1016/j.ecns.2010.05.005

Larsen, D. P., Butler, A. C., & Roediger Iii, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, *42*(10), 959-966. https://doi.org/10.1111/j.1365-2923.2008.03124.x

Min Simpkins, A. A., Koch, B., Spear-Ellinwood, K., & St. John, P. (2019). A developmental assessment of clinical reasoning in preclinical medical education. *Medical Education Online*, *24*(1), 1591257. https://doi.org/10.1080/10872981.2019.1591257

O'Brien, S. R., Dillon, N., Linsky, M., Lagueras, K., Uhl, J., Conroy, S., Kiley, A., Lanzafame, N., States, E., & Sturm, E. (2023). Initial validation of a script concordance test to measure the development of clinical reasoning among physical therapy residents. *The Journal of Clinical Education in Physical Therapy*, *5*. https://doi.org/10.52214/jcept.v5.9014

Pinnock, R., Young, L., Spence, F., Henning, M., & Hazell, W. (2015). Can Think Aloud Be Used to Teach and Assess Clinical Reasoning in Graduate Medical Education? *Journal of Graduate Medical Education*, *7*(3), 334-337. https://doi.org/10.4300/JGME-D-14-00601.1

Qiao, J., Huang, C.-R., Liu, Q., Li, S.-Y., Xu, J., & Ouyang, Y.-Q. (2023, Month). Effectiveness of NonImmersive Virtual Reality Simulation in Learning Knowledge and Skills for Nursing Students: Metaanalysis. Clinical Simulation in Nursing, 76, 26-38. https://doi.org/10.1016/j.ecns.2022.12.003.

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rushton, D. A., Beeton, D. K., Jordaan, D. R., Langendoen, M. J., Levesque, M. L., & Maffey, M. L. (2016). PART A: EDUCATIONAL STANDARDS 2016. *International Federetion for Orthopeadics Manipulative Therapists.*

Schuwirth, L. (2009). Is assessment of clinical reasoning still the Holy Grail? *Medical Education*, *43*(4), 298-300. https://doi.org/10.1111/j.1365-2923.2009.03290.x

Smart, K., & Doody, C. (2007). The clinical reasoning of pain by experienced musculoskeletal physiotherapists. *Manual Therapy*, *12*(1), 40-49. https://doi.org/10.1016/j.math.2006.02.006

Thammasitboon, T., Jj, R., Rl, T., Apj, O., M, S., & G, D. (2018). The Assessment of Reasoning Tool (ART) : Structuring the conversation between teachers and learners. *Diagnosis (Berlin, Germany)*, *5*(4). https://doi.org/10.1515/dx-2018-0052

Tubelo, R. A., Portella, F. F., Gelain, M. A., De Oliveira, M. M. C., De Oliveira, A. E. F., Dahmer, A., &

Pinto, M. E. B. (2019). Serious game is an effective learning method for primary health care education of medical students: A randomized controlled trial. *International Journal of Medical Informatics*, *130*, 103944. https://doi.org/10.1016/j.ijmedinf.2019.08.004

Young, M., Thomas, A., Gordon, D., Gruppen, L., Lubarsky, S., Rencic, J., Ballard, T., Holmboe, E., Da Silva, A., Ratcliffe, T., Schuwirth, L., & Durning, S. J. (2019). The terminology of clinical reasoning in health professions education: Implications and considerations. *Medical Teacher*, *41*(11), 1277-1284. https://doi.org/10.1080/0142159X.2019.1635686

Viau, R. (2009). La motivation en contexte scolaire (5e éd.). De Boeck.

Appendix: Numerical values of the results

Table A1. Numerical results obtained in the case where items were randomly removed. Average scores (of 20 generated games) are given with their standard deviation. The last two columns give the minimal and maximal scores that were generated.

Error (%)	Average Score	Score standard deviation	Min	Max
0	100	0	100	100
5	97.80	2.07	93	100
10	93.30	2.47	88	97
15	90.95	5.38	76	99
20	84.95	7.54	65	94
25	82.65	6.70	63	91
30	77.65	7.88	63	93
35	72.45	9.26	53	89
40	69.50	9.49	49	88
45	58.80	11.28	27	75
50	50.75	11.65	33	75
55	48.40	14.20	25	73
60	36.85	11.54	18	60
65	34.25	8.42	21	47
70	26.40	7.69	10	43
75	21.50	4.49	15	33
80	20	6.01	10	35
85	15.95	4.33	10	25

90	11	4.42	0	16
95	5.50	3.10	0	8
100	0	0	0	0

Table A2. Same as Table A1 in the case where items were randomly replaced.

Error (%)	Average Score	Score standard deviation	Min	Max
0	100	0	100	100
5	94.25	2.71	89	99
10	89.80	4.94	80	98
15	83.15	6.83	69	95
20	76.05	7.97	65	89
25	69.70	8.77	50	86
30	67.60	6.68	53	80
35	60.45	9.83	36	76
40	54.75	8.43	38	67
45	45.00	9.88	26	62
50	39.30	10.42	23	63
55	35.45	8.99	21	53
60	30.35	10.20	13	53
65	28.15	11.88	12	62
70	20.60	10.42	6	45
75	15.35	8.59	0	38
80	9.10	5.25	0	23
85	11.60	7.35	0	25

90	3.35	6.11	0	24
95	2.85	5.21	0	20
100	0.40	1.19	0	5