



Evaluation expérimentale du raisonnement clinique dans le cadre des jeux sérieux pour la formation des professionnels de santé.

Antonia Blanié Blanié

► To cite this version:

Antonia Blanié Blanié. Evaluation expérimentale du raisonnement clinique dans le cadre des jeux sérieux pour la formation des professionnels de santé.. Neurosciences [q-bio.NC]. Université Paris-Saclay, 2020. Français. NNT : 2020UPASS092 . tel-03245702

HAL Id: tel-03245702

<https://theses.hal.science/tel-03245702>

Submitted on 2 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Évaluation expérimentale du raisonnement clinique dans le cadre des jeux sérieux pour la formation des professionnels de santé

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 566 -Sciences du sport, de la motricité et du
mouvement Humain (SSMMH)

Spécialité de doctorat : Psychologie

Unité de recherche : -Complexité, Innovation, Activités Motrices et
Sportives (CIAMS)

Référent : Faculté des Sciences du Sport

Thèse présentée et soutenue au Kremlin-Bicêtre,
le 19 juin 2020, par

Antonia Blanié

Composition du Jury

Olivier LAMBOTTE

Professeur, Université Paris Saclay

Domitile LOURDEAUX

Maitre de conférences, HDR, UTC

Patrick PLAISANCE

Professeur, Université de Paris

Corinne LEJUS-BOURDEAU

Professeur, Université de Nantes

Michel-Ange AMORIM

Professeur, Université Paris Saclay

Dan BENHAMOU

Professeur, Université Paris Saclay

Président

Rapporteur & Examinatrice

Rapporteur & Examineur

Examinatrice

Directeur de thèse

Co-Directeur de thèse

Titre : Évaluation expérimentale du raisonnement clinique dans le cadre des jeux sérieux pour la formation des professionnels de santé

Mots clés : Jeux sérieux, simulation, formation, professionnels de santé, raisonnement clinique

Résumé : L'amélioration du raisonnement clinique (RC) est un enjeu essentiel pour la Médecine de demain car il est établi que son utilisation imparfaite conduit à des résultats de soins insuffisants. Le RC est un processus cognitif complexe. Cette activité intellectuelle synthétise l'information obtenue à partir de la situation clinique et l'utilise pour faire une analyse diagnostique et prendre une décision de prise en charge du patient en intégrant les connaissances et expériences antérieures. La formation à cette compétence est donc essentielle. Pour améliorer le raisonnement, une connaissance des mécanismes qui le constituent est nécessaire et une revue de ces mécanismes constitue la partie initiale de cette thèse.

La formation des professionnels de santé par la simulation se généralise avec pour objectif « jamais la première fois sur le patient ». En plein essor, le jeu sérieux (JS) représente un outil pédagogique intéressant. Une revue de la littérature sur l'efficacité des JS et plus particulièrement dans le cadre du RC est également incluse dans la partie initiale de cette thèse. Ainsi, les JS sont efficaces et peuvent, entre autres, cibler certaines compétences dont le RC. Cependant, la plupart des études sur le RC dans le cadre des JS, restent subjectives avec des évaluations qualitatives ou des autoévaluations des apprenants ou concernent uniquement le résultat (prise de décision). Par conséquent, la valeur éducative et les modalités des JS dans la formation du RC restent à approfondir.

Le JS *LabForGames Warning* a été développé dans le centre de simulation LabForSIMS pour les étudiants infirmiers et cible la détection de l'aggravation d'un patient et la communication. Cette thèse a pour objectif de tester un mode d'apprentissage utilisant la simulation par les JS afin d'améliorer le RC chez les professionnels de santé.

Une 1ère étude a évalué la validité du JS *LabForGames Warning* selon le cadre théorique proposé par Messick. Elle a démontré que les scores et le temps de jeu ne pouvaient pas différencier le niveau des compétences cliniques des infirmiers.

Cependant, les preuves de validité étaient obtenues pour le contenu, le processus de réponse et la structure interne. Même si cette version du jeu ne peut donc pas être utilisée pour une évaluation sommative des étudiants, notre étude montre que ce JS est bien accepté par les étudiants et qu'il peut être utilisé pour la formation au sein d'un programme éducatif. Une seconde étude a évalué l'efficacité de 2 modalités pédagogiques sur l'apprentissage du RC à la détection de l'aggravation clinique d'un patient en comparant un groupe d'étudiants infirmiers formé par simulation avec *LabForGames Warning* par rapport à un groupe formé par un enseignement traditionnel. Le RC a été évalué par les tests de concordance de script immédiatement et 1 mois après. Cette étude multicentrique randomisée a inclus 146 étudiants infirmiers volontaires. Aucune différence significative n'a été observée sur le RC entre la formation par la simulation avec JS et l'enseignement traditionnel. Cependant, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation.

En conclusion, nous avons tout d'abord confirmé la validité du JS *LabForGames Warning* en tant qu'outil pédagogique à visée formative et non sommative. Puis, bien qu'aucune différence d'apprentissage du RC n'ait été observée entre la formation par la simulation avec JS et l'enseignement traditionnel, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation avec le jeu. Les études sont à poursuivre pour préciser les modalités et stratégies pédagogiques des JS dans la formation des professionnels de santé, comme par exemple la place du débriefing, le rôle de la motivation. En effet, en plein essor, les développements technologiques telle que l'intelligence artificielle vont transformer la formation au RC ainsi que les outils pédagogiques disponibles dans les années à venir.

Title : Experimental Evaluation of Clinical Reasoning in Serious Games for the Education of Health Professionals

Keywords : Serious games, simulation, training, healthcare professionals, clinical reasoning

Abstract : Improvement of clinical reasoning (CR) is a key issue for the future of medicine because it has been established that imperfect reasoning leads to insufficient care results. CR is a complex cognitive process which synthesizes information obtained from the clinical situation, then uses it to make a diagnostic analysis and take a decision on patient management by integrating previous knowledge and experience. Training for this skill is therefore essential. To improve reasoning, knowledge of the mechanisms which build it up is necessary and a review of these mechanisms constitutes the initial part of this thesis. Training of healthcare professionals through simulation is becoming widespread with the objective of "never the first time on a patient". In rapid expansion, the use of serious games (SG) represents an interesting pedagogical tool. A review of the literature on the effectiveness of SG and more particularly in the context of CR is also included in the initial part of this thesis. Thus, SG is effective and may, among other things, target certain skills, including CR. However, most studies dealing with CR by using SG include qualitative assessments or self-assessments of learners or focus only on the outcome (decision making). Therefore, the educational value and modalities of SG in the training of CR of health professionals remain to be further explored.

The SG *LabForGames Warning* was developed in the LabForSIMS simulation center for nursing students and targets detection of patient deterioration and the ensuing communication. The objective of this thesis is to test a learning mode using SG simulation in order to improve CR in healthcare professionals.

The first study evaluated the validity of the SG *LabForGames Warning* according to the Messick's Framework. This study showed that scores and playing time could not differentiate the level of clinical skills of nurses. However, evidence of validity was obtained for content, response process and internal structure. Although this version of the game cannot therefore be used for summative evaluation of students, our study shows that this SG is well accepted by students and can be used for training within an educational program.

A second study evaluated the effectiveness of 2 teaching modalities on learning CR for the detection of clinical patient deterioration by comparing a group of nursing students trained by simulation with *LabForGames Warning* compared to a group trained by traditional teaching. CR was assessed by script concordance tests immediately and 1 month later. This randomized multicenter study included 146 volunteer nursing students. No significant difference was observed in CR change between simulation training with SG and traditional teaching. However, satisfaction and motivation were better with simulation instruction.

In conclusion, we have confirmed the validity of SG *LabForGames Warning* as an educational tool with formative and not summative aims. Then, although no difference in learning about CR was observed between simulation training with SG and traditional teaching, satisfaction and motivation were better with simulation teaching with the game. Further studies are needed to clarify the modalities and pedagogical strategies of SG in the training of healthcare professionals, such as the place of debriefing and the role of motivation. Moreover, technological developments such as artificial intelligence might transform CR training and the available pedagogical tools in the coming years

Table des matières

Mes remerciements,	6
Avant-propos	7
I - Revue de la littérature scientifique	9
1. Le raisonnement clinique en santé	9
1.1. Définitions : Raisonnement clinique et prise de décision	10
1.2. Modèles théoriques de prise de décision – raisonnement clinique	11
1.3. Méthodes d'évaluation du raisonnement clinique	13
2. Les jeux sérieux en santé	21
2.1 Définitions	21
2.2 Jeux sérieux et théories d'apprentissage	24
2.3 Efficacité des jeux sérieux dans le processus d'apprentissage	25
2.4. Le raisonnement clinique dans les jeux sérieux	28
II - Contribution expérimentale	37
1. Problématique et hypothèses	37
2. Développement du jeu sérieux <i>LabForGames Warning</i>	38
3. Étude 1: Évaluation de la validité du jeu sérieux <i>LabForGames Warning</i> .	43
3.1. Introduction : les cadres théoriques de validité d'un outil pédagogique	43
3.2. Article 1- Assessing validity evidence for a serious game dedicated to patient clinical deterioration and communication	47
3.3. Conclusion	74
4. Étude 2 : Évaluation de l'efficacité du jeu sérieux sur l'apprentissage du raisonnement clinique nécessaire à la détection de l'aggravation d'un patient chez les étudiants infirmiers en comparant une simulation avec jeu sérieux et un enseignement traditionnel.	75
4.1. Introduction : raisonnement clinique et détérioration de l'aggravation clinique d'un patient	75
4.2. Article 2: 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 ¹²⁴ .	75
4.3. Conclusion	87
III- Discussion Générale	88
1. Motivation et jeux sérieux	88
2. Jeux sérieux, raisonnement clinique, et stratégies pédagogiques	89
3. Raisonnement clinique et intelligence artificielle	91

4. Apport du débriefing dans le jeu sérieux	92
5. Le jeu sérieux du futur	93
Conclusion	93
Références	94
Annexe 1 : Curriculum Vitae	102

Mes remerciements,

s'adressent en premier lieu à monsieur le Professeur Benhamou à qui ces quelques mots ne suffiront pas à exprimer toute ma gratitude. Vous m'avez fait confiance depuis le début de ma carrière hospitalière. Vous m'avez poussé à entreprendre et à contribuer à bon nombre de projets cliniques et universitaires. Travailler à vos côtés au Centre hospitalier universitaire de Bicêtre depuis déjà 10 ans, est une grande chance et me permet, chaque jour, d'approfondir mes connaissances et ma réflexion dans l'anesthésie, la recherche et la pédagogie.

A monsieur le Professeur Amorim, qui a accepté de codiriger ce travail et m'a accordé sa confiance. Votre encadrement, votre rigueur et vos conseils m'ont guidé tout le long de ces quatre années.

A monsieur le Professeur Lambotte, qui me fait l'honneur de présider le jury de cette thèse. Ce travail m'a permis de poursuivre mes recherches dans le domaine passionnant de la pédagogie.

A madame Lourdeaux et monsieur le Professeur Plaisance, qui me font l'honneur d'être rapporteurs de cette thèse. Veuillez recevoir toute ma reconnaissance pour le temps accordé à ce travail, en particulier dans les circonstances exceptionnelles liées à la pandémie que nous connaissons actuellement.

A madame le professeur Lejus-Bourdeau, qui me fait l'honneur de juger cette thèse. Veuillez recevoir tous mes remerciements et l'expression de mon profond respect.

A monsieur le Professeur Duranteau, que je souhaiterais remercier pour son soutien et de m'avoir permis de mener à bien ces travaux. Votre confiance, votre écoute et vos conseils me sont très précieux dans tous mes projets au sein du service.

A mes collègues Bicestrais, à cette belle équipe d'anesthésie-réanimation, à laquelle je suis fière d'appartenir. C'est grâce à vous que j'ai pu accomplir tous mes projets. Un remerciement à Philippe, Christine, Laurent, Dimitri, Hanane et bien d'autres...

A mes parents qui m'ont toujours soutenu dans toutes les étapes de ma vie. Vous êtes des modèles pour moi, me montrant que l'épanouissement familial et professionnel sont totalement compatibles.

A Sébastien, qui, à mes côtés, a su être patient avec moi. Grâce à toi, j'ai découvert le plaisir de me laisser porter... c'est tellement apaisant.

A Elisa, mon petit cœur, à qui j'ai hâte de faire très vite de vrais gros câlins. Tu as su m'écrire de si jolis mots de soutien sur tes dessins, ils me font chaud au cœur.

Et bien sûr à ma petite famille recomposée, et à tous mes amis qui ont su être présents tout le long de cette thèse.

Avant-propos

Médecin anesthésiste réanimateur depuis 2010, j'ai eu une activité initialement mixte en anesthésie et en réanimation puis principalement en anesthésie au CHU de Bicêtre. Les blocs opératoires du CHU sont polyvalents avec de la chirurgie majeure, des polytraumatisés, la grande garde de neurochirurgie mais aussi de la chirurgie ambulatoire. En plus de ma participation à l'ensemble de ces activités, je suis, depuis 2019, coordonnateur adjoint de l'unité de chirurgie ambulatoire. Par ailleurs, mon intérêt pour la qualité et la gestion des risques m'a conduit, depuis 2017, à être membre du Comité Analyse et Maîtrise du Risque de la Société Française d'Anesthésie et de Réanimation (SFAR).

En parallèle, je me suis intéressée à la formation des professionnels de santé et à la simulation en particulier. J'ai développé une expertise dans ce domaine avec un DU de formateur en simulation en 2013 et la participation à de nombreux congrès et formations dans le domaine de la pédagogie et de la simulation. J'ai également contribué au développement du centre de simulation de la Faculté de Médecine Paris Sud LabForSIMS en tant que responsable adjointe. Enfin, je participe activement à la SoFraSimS (Société Francophone de Simulation en Santé) comme membre du comité recherche et membre du conseil d'administration. Un curriculum Vitae détaillé est disponible en annexe.

Ainsi, en tant que formateur en simulation en santé et médecin anesthésiste réanimateur, ce doctorat m'a permis d'approfondir mes compétences dans le domaine de la simulation et pédagogie en santé, et plus particulièrement l'aspect pédagogique des jeux sérieux. Le projet de recherche était dans la continuité des thématiques de recherche du centre de simulation LabForSIMS (rattaché au CIAMS). En effet, l'équipe de LabForSIMS a construit deux jeux sérieux (dont j'ai coordonné les projets) qui ont été parmi les premiers disponibles en langue française. Un premier jeu sérieux pour les étudiants en médecine cible les pathologies neurologiques aiguës aux urgences (*LabForGames Neuro*) (APP innovation pédagogique Université Paris Sud 2014). Il est utilisé avec les étudiants en médecine lors d'enseignements dirigés. Un second pour les étudiants en soins infirmiers cible la détection de l'aggravation d'un patient et la communication interprofessionnelle avec l'outil SAED (*LabForGames Warning*) (APP ARS simulation en santé 2015). Alors que ces jeux sérieux sont un moyen reconnu et accepté de formation, très apprécié des étudiants, de nombreuses inconnues persistent quant à leur valeur pédagogique. Les mécanismes cognitifs mis en jeu lors de l'apprentissage avec ces jeux sérieux ne sont pas clairement déterminés et sont explorés par des travaux de recherche du laboratoire. Ces recherches sont centrées sur l'apprentissage du raisonnement clinique, la prise de décision, la communication, mais aussi la détection de l'aggravation clinique qui sont des compétences essentielles pour les professionnels de santé et plus précisément dans ma spécialité en anesthésie réanimation. La crise sanitaire actuelle avec la prise en charge de patients Covid19 positifs en est un exemple. En parallèle à cette thèse, le projet ANR « LabForSIMS2 » sur la création d'un environnement virtuel collaboratif et immersif m'a permis de

participer à un projet de recherche avec le développement de briques technologiques telles que l'analyse conversationnelle et la réalité mixte dans des scénarios de réanimation néonatale et de prise en charge d'urgence chirurgicale.

Sur le long terme, ce travail va me permettre de poursuivre le développement du centre de simulation de la Faculté de médecine Paris Saclay, et plus particulièrement, de développer des projets de recherche dans ce domaine. Ces projets donneront lieu à des publications. Ces compétences seront également mises à profit dans la Commission d'innovation pédagogique de la Faculté de Médecine Paris Saclay à laquelle je participe.

I - Revue de la littérature scientifique

1. Le raisonnement clinique en santé

Le soin clinique de qualité nécessite une succession d'analyses et de décisions qui constituent le raisonnement clinique (diagnostique, demande d'examen complémentaires, thérapeutique, décision sur le parcours de soins, suivi et ré-évaluation, etc.). Jusqu'à récemment, la question des erreurs médicales au niveau du diagnostic étaient peu mises en avant. Depuis peu, ce sujet est apparu comme une question de santé publique, tant en soins primaires^{1,2} qu'en milieu hospitalier, tel que la réanimation³. En médecine de ville, on estime qu'une erreur diagnostique survient chez environ 5 % des patients et que la moitié d'entre elles conduisent à un dommage². Pour l'OMS, la réduction du taux d'erreurs diagnostiques est une haute priorité^{1,4}. L'incidence des erreurs de diagnostic varie selon les spécialités médicales : entre 2 à 12% pour les diagnostics radiologiques et 12 à 15% pour la médecine d'urgence³. En réanimation pédiatrique et néonatale, jusqu'à 20 % des diagnostics s'avèrent faux à l'autopsie⁵.

Les approches pour remédier à cette question nécessitent d'en comprendre les mécanismes. Ceux-ci s'avèrent multiples et sont décrits soit sous forme de recommandations d'amélioration des pratiques (améliorer l'accès aux tests diagnostiques, améliorer les technologies de l'information, augmenter le retour d'information aux cliniciens sur leurs diagnostics...)¹, soit sous forme de stratégies pragmatiques telles que l'amélioration de la communication et la création de checklists⁶, d'algorithmes, ou de revues de cas⁷.

Parmi les causes des erreurs diagnostiques, il est apparu qu'environ 75 % des erreurs diagnostiques sont liées à un défaut du processus cognitif du clinicien⁸. Les difficultés de raisonnement clinique affectent entre 5 et 15 % des étudiants en médecine et il est souvent détecté tardivement⁹. Le raisonnement clinique, au cœur de la pratique médicale, est ainsi un véritable challenge pédagogique. Il s'agit donc bien d'un mécanisme essentiel qu'il faut non seulement comprendre mais sur lequel il semble nécessaire d'agir^{10,11}. Des stratégies de remédiation ont été proposées par Audétat et al. pour chaque difficulté spécifique du processus de raisonnement clinique¹². Bien que certaines études aient tenté des actions correctrices, les approches cognitives sont encore rares, peu développées et les méthodologies encore imparfaites^{11,13}. Parmi celles-ci, les interventions éducatives, c'est-à-dire qui ont pour but de modifier en profondeur le raisonnement clinique, sont peu développées et leurs résultats inconstants. De plus, les méthodes de pédagogie innovante n'ont été que peu utilisées et c'est notamment le cas de la simulation et de l'emploi des méthodologies numériques¹⁴⁻¹⁶.

Cette thèse a pour but de tester un mode d'apprentissage utilisant la simulation virtuelle par les jeux sérieux afin d'améliorer le raisonnement clinique. Pour l'améliorer,

une connaissance des mécanismes qui le constituent est nécessaire et une revue de ces mécanismes constitue la partie initiale de cette thèse.

1.1. Définitions : Raisonnement clinique et prise de décision

Le raisonnement clinique est un processus cognitif complexe. Il correspond à l'ensemble des processus cognitifs permettant aux professionnels de santé d'évaluer et de gérer les problèmes de santé d'un patient^{9,17,18}. Cette activité intellectuelle synthétise l'information obtenue à partir de la situation clinique et l'utilise pour faire une analyse diagnostique et prendre une décision de prise en charge du patient en intégrant les connaissances et expériences antérieures. Reasonner permet ainsi de poser des diagnostics et de prendre des décisions.

Le raisonnement clinique est une compétence centrale en médecine et concerne tous les professionnels de santé^{18,19}. Levett-Jones et Lee^{20,21} ont par exemple décrit les différentes étapes successives du processus du raisonnement clinique infirmier : L'infirmier/ère va recueillir des indices, traiter l'information, comprendre un problème ou une situation du patient, planifier et mettre en œuvre des interventions, évaluer les résultats et réfléchir au processus et en tirer des leçons (Figure 1). Le raisonnement clinique est donc un processus cyclique.

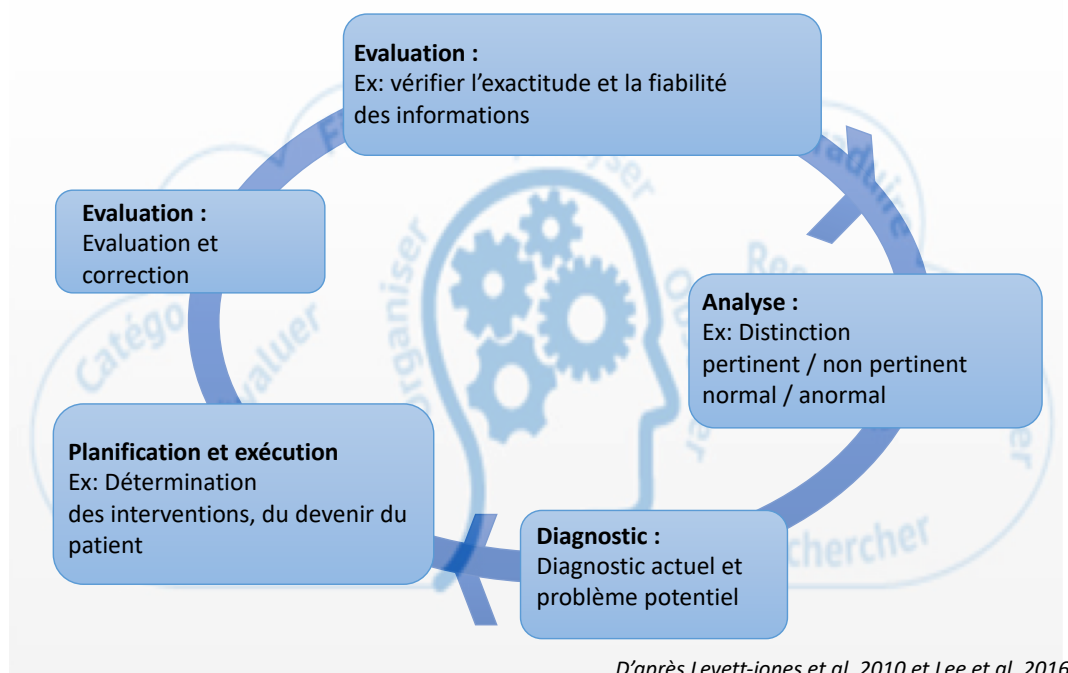


Figure 1 : Processus de description du raisonnement clinique chez les infirmiers (d'après Levett-Jones 2010²⁰ et Lee 2016²¹).

La prise de décision est souvent assimilée au raisonnement clinique dans la littérature puisqu'elle en est une étape^{20 18}. Elle est définie comme un processus cognitif aboutissant à un jugement et permettant de décider des mesures appropriées pour

répondre aux besoins d'une situation clinique donnée. Concernant les compétences non techniques décrites par Fletcher et Gaba, la prise de décision fait partie de la catégorie des compétences cognitives au même titre que la conscience de la situation (Tableau 1) ^{22,23}.

Tableau 1: Compétences non techniques décrites par Fletcher et Gaba ^{22,23}.

Catégorie des compétences non techniques	
Compétences cognitives	Conscience de la situation
	Prise de décision
Compétences comportementales ou sociales	Travail en équipe
	Communication
	Leadership/ Followership
	Assertivité
Ressources personnelles	Gestion de son stress
	Gestion de sa fatigue

1.2. Modèles théoriques de prise de décision – raisonnement clinique

Différents modèles théoriques expliquant le raisonnement clinique et la prise de décision ont été décrits dans la littérature ^{3,8,17,24-26}.

1.2.1. Raisonnement heuristique ²⁴

Décrit par Kahneman, le raisonnement heuristique correspond à un raisonnement avec l'utilisation de raccourcis mentaux pour diminuer la charge cognitive de la prise de décision²⁴. Cela correspond à une reconnaissance de scripts cliniques (*illness script ou pattern*). En médecine, plusieurs types d'heuristiques ont été décrits. L'heuristique de représentativité est une stratégie de décision basée sur une ressemblance à modèle mental (plus qu'une probabilité statistique). L'heuristique de disponibilité est la tendance à penser à un événement en mémoire plutôt qu'avec sa fréquence. Le 3ème heuristique est l'heuristique d'ancrage (*anchoring*) qui est une impossibilité à modifier un diagnostic ou une décision initiale malgré de nouvelles informations.

1.2.2. Théorie du double processus (*Dual processus Theory*)

Un des plus connus est celui de la théorie du double processus (Figure 2)^{17,19,25}. Cette théorie hybride utilise 2 modèles séparés et suppose le passage de l'un à l'autre.

Le modèle du **raisonnement intuitif** (Système 1) est un modèle de raisonnement heuristique, automatique. L'opération mentale est rapide et intuitive. Ce système réflexe est lié à des modèles de reconnaissance (*pattern recognition*) ou de scripts cliniques. Il est fortement lié à l'expérience.

Le modèle du **raisonnement analytique** (Système 2) correspond à une approche hypothético-déductif. C'est une analyse délibérée et consciente qui nécessite un effort

et du temps. Ce modèle est souvent utilisé dans des situations : complexes, où il y a le temps, avec un contexte certain ou une décision hautement risquée. Ce système est peu efficient dans les situations d'urgence.

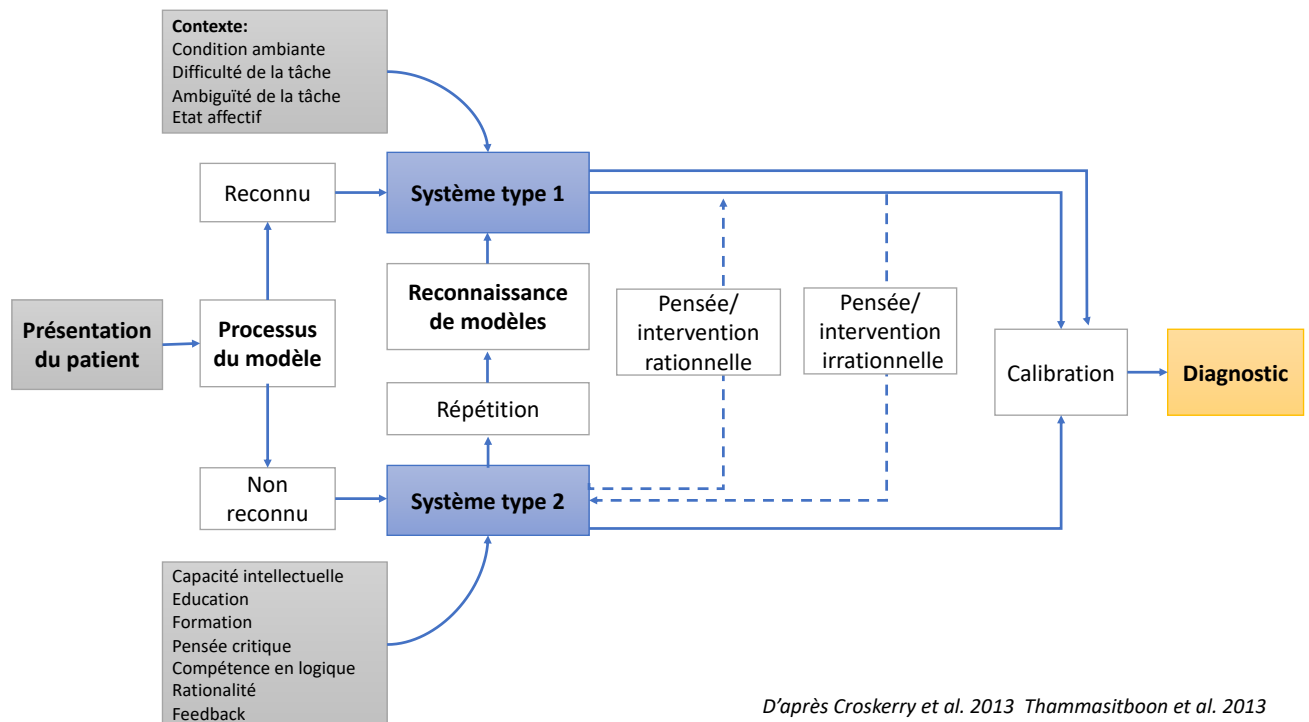


Figure 2: Modèle de la Théorie du double processus (*Dual process Theory*). D'après Croskerry et al. et Thammasitboon et al. ^{8,25}

Cependant, on ne peut pas généraliser :

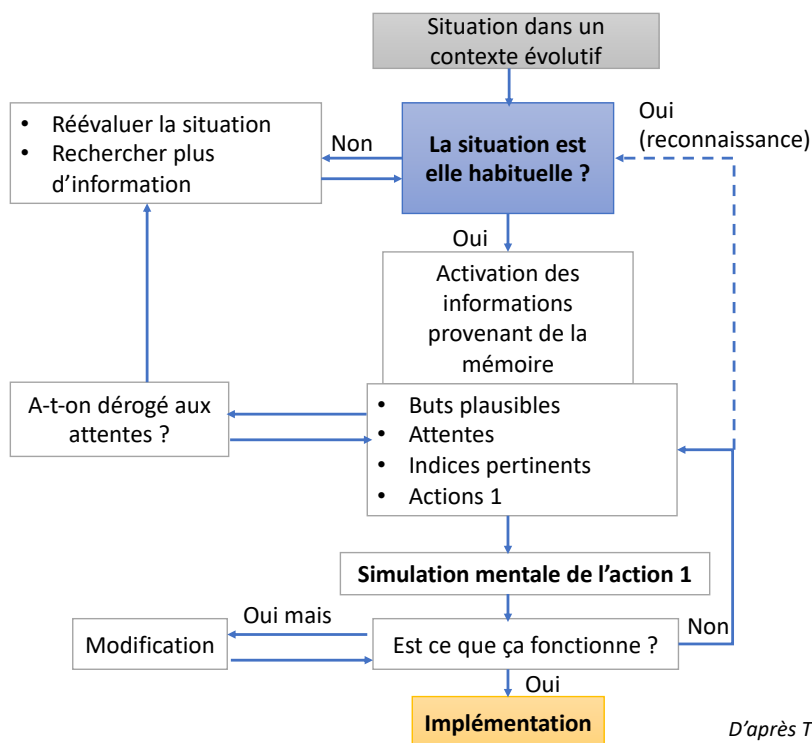
- un cas simple/standard comme utilisant un modèle heuristique et un cas complexe comme analytique.
- ni un expert comme utilisant un modèle heuristique et un novice un modèle analytique

Le professionnel de santé utilise en pratique les deux systèmes et passe d'un système à l'autre en fonction du contexte et de la situation.

Ces modèles sont basés sur le fait que notre système cognitif conscient a une capacité limitée de traitement, alors que notre système cognitif non-conscient est caractérisé par beaucoup de routine ou automaticité. L'intuition est donc indispensable dans le raisonnement clinique¹⁷. Elle est permise par la présence de scripts cliniques ou de modèles de reconnaissance qui sont stockés dans la mémoire à long terme. Par exemple, avant même de voir le patient, le clinicien urgentiste émet des hypothèses diagnostiques. Ainsi, dans les premières minutes d'un entretien, le clinicien a déjà élaboré plus de 75% des hypothèses²⁷. Cependant l'intuition est un processus subjectif influencé par de nombreux facteurs : l'attention, la disponibilité, la mémorisation, les expériences antérieures, et d'autres signaux extérieurs.

1.2.3. Modèle *recognition-primed decision-making*

Ce modèle naturaliste décrit par Klein²⁸, consiste à reconnaître un modèle clinique typique, comme précédemment décrit, et d'y appliquer la stratégie typique associée (Figure 3)⁸. Mais avant de l'appliquer, une étape d'analyse est réalisée grâce à une simulation mentale cherchant à décrire comment la solution imaginée pourrait fonctionner dans le contexte de la situation actuelle.



D'après Thammasitboon et al. 2013

Figure 3 : Modèle de *recognition-primed decision-making*⁸.

1.3. Méthodes d'évaluation du raisonnement clinique

1.3.1. Évaluation de la prise de décision.

Le raisonnement clinique peut être évalué en mesurant le résultat de ce raisonnement, c'est à dire la prise de décision : le bon diagnostic, le bon traitement... Par exemple, dans le cas virtuel du jeu sérieux « ALICE », le raisonnement clinique est évalué par les auteurs par la prise de décision: bon traitement ou bon diagnostic²⁹.

Cependant, la limite majeure de cette évaluation réside dans le fait que les décisions médicales peuvent en fait être classées en :

- bonne action pour une bonne raison
- bonne action pour une mauvaise raison
- mauvaise action pour une bonne raison
- mauvaise action pour une mauvaise raison

Il est donc important d'évaluer tout le processus du raisonnement clinique et non uniquement l'étape finale³.

1.3.2. Analyse qualitative

Le raisonnement clinique peut être évalué grâce une analyse qualitative. Une première méthode est la méthode de penser à haute voix (« *think-aloud method* »). Il est demandé aux apprenants de penser à haute voix, et leurs données verbales sont ensuite analysées^{21,30}. L'intérêt par rapport à l'observation seule est que cela permet de lier ce qui est fait à un processus cognitif. Cependant ; on ne leur demande pas précisément d'expliquer leurs processus de pensée (raisonnement) et le fait de penser à haute voix peut parfois interrompre des stratégies cognitives. Une seconde méthode est la réalisation d'interviews rétrospectifs permettant d'analyser le processus de décision et de raisonnement mais rétrospectivement ce qui entraîne d'autres biais. Pour évaluer la conscience de la situation, on peut utiliser la méthode SAGAT (*situation awareness global assessment technique*)³¹. Dans cette méthode, le scénario clinique est interrompu plusieurs fois au cours duquel le participant va répondre à des questions clés concernant : (1) la perception des informations, (2) la compréhension de leur signification et (3) la projection. Les réponses permettent d'en évaluer la conscience de la situation.

Ainsi, l'analyse qualitative est intéressante mais a des limites rendant des évaluations quantitatives nécessaires.

1.3.3. Les questions à choix multiples (QCM).

Les questions à choix multiples (QCM), bien que souvent utilisées, mesurent imparfaitement la qualité du raisonnement clinique³². Elles mesurent plutôt la connaissance. Cependant ; elles possèdent une excellente valeur prédictive en termes de performance sur la pratique future³².

1.3.4. L'examen clinique objectif et structuré (ECOS)

L'examen clinique objectif et structuré (ECOS) évalue la démarche clinique par observation directe à partir de situations cliniques simulées (vrais patients, acteurs, simulateur basse fidélité...) et standardisées (tous les candidats sont soumis aux mêmes tâches cliniques)³². L'ECOS comporte des situations multiples, de durée prédéterminée, qui évaluent chacune un comportement ou une performance distincte. L'évaluateur utilise une grille d'observation prédéfinie (checklist) avec une bonne fidélité inter-juges. L'ECOS est habituellement utilisé pour évaluer des compétences cliniques. Sa validité est élevée puisqu'il existe une corrélation entre le score à l'ECOS et la qualité de la pratique en situation réelle³³. Mais il est moins valide pour l'évaluation du raisonnement clinique car, de par sa nature, il n'évalue que des comportements observables³². Or, il est difficile de mesurer les subtilités d'une démarche de raisonnement même avec des grilles et même exprimée à haute voix. Les grilles conduisent donc plus souvent à récompenser la minutie dans le recueil de données que véritablement les qualités du raisonnement.

1.3.5. Les grilles globales d'évaluation

Les grilles globales d'évaluation telles que les checklists comportent plusieurs items construits par des experts. Ces grilles sont ensuite remplies par un ou plusieurs observateurs en observant le comportement de l'apprenant lors d'un stage ou d'une simulation. La multiplicité des critères permet d'évaluer un grand nombre de compétences cliniques mais la plupart du temps l'évaluation du raisonnement clinique ne représente qu'une partie de la grille. Par exemple, l'équipe de Liaw et al. a développé une échelle évaluant la performance sur la détection et la prise en charge de la détérioration clinique d'un patient pour les infirmiers³⁴. Cette grille (RAPIDS) a été construite par des experts et comportait 42 items. La plupart des items testaient plutôt des performances à réaliser des compétences cliniques (« initier de l'oxygène », « faire une glycémie »...) mais quelques items évaluaient un peu plus le raisonnement (« donner l'évaluation du problème »). Cette grille avait été validée par cette équipe sur le plan du contenu mais aussi concernant sa validité de construit du score (adéquation entre les scores obtenus et le degré d'expertise) et sa fiabilité inter observateurs³⁴.

Il existe des grilles encore plus globales (« *global rating scale* ») mais qui évaluent de manière unique une compétence, sans le détail des items, en utilisant en général une échelle de Likert. Par exemple, l'étude de Liaw et al. ³⁴ utilisait également deux « *global rating scale* » avec une échelle de Likert de 1 à 10: une sur « l'évaluation et la prise en charge de la détérioration clinique » et l'autre sur « la transmission de la détérioration clinique »³⁴. Les auteurs ont d'ailleurs trouvé une bonne corrélation entre le score obtenu avec la checklist et celui de la *global rating scale* ($r=0,94$, $p<0,001$). Ainsi, comme pour les ECOS, ces grilles utilisent des items avec des éléments observables de compétences cliniques et n'évaluent donc pas réellement la démarche de raisonnement elle-même.

1.3.6. Les problèmes de gestion de patient (« *patient management problem* »)

Le « *patient management problem* » correspond à reproduire sur papier ou ordinateur la prise en charge habituelle d'un patient étape par étape c'est-à-dire selon le processus avec lequel un médecin collecte les informations par l'interrogatoire et par l'examen clinique, puis prend les décisions liées au diagnostic, à l'investigation ou à la prise en charge thérapeutique³². Les différents éléments sont dévoilés au fur et à mesure que l'apprenant avance dans sa démarche. Le parcours suivi par l'apprenant est ensuite comparé à celui d'un expert et des scores lui sont attribués. Ce type d'évaluation a été beaucoup utilisé aux États-Unis mais plusieurs limites existent. Tout d'abord, l'effet « indice » consiste en ce que le candidat peut reconnaître la bonne réponse plutôt que la générer lui-même, ce qui fausse la compétence de résolution de problème et améliore artificiellement la performance. De plus, le cheminement idéal n'existe pas et les experts utilisent des cheminements de résolution de problème qui peuvent être différents entre chaque expert. Ensuite, l'exhaustivité et la minutie de la prise en charge du patient entraînent un bon score mais ne sont pas forcément le reflet de la performance clinique en raisonnement. En effet, les experts utilisent souvent des

raccourcis via leur reconnaissance de scripts cliniques et peuvent alors avoir de moins bons scores que des novices qui sont exhaustifs mais ont un moins bon raisonnement. De plus, la compétence à résoudre un problème clinique ne permet pas de prédire avec certitude la capacité à résoudre un autre problème. L'évaluation utilisant un long cas unique comme dans le « *patient management problem* » n'est pas conseillée quand on cherche à évaluer le processus de raisonnement clinique. Il est plutôt conseillé d'utiliser plusieurs évaluations de plusieurs cas courts³².

1.3.7. Les problèmes à éléments clés (« key-feature problems », KFP)

Les problèmes à éléments clés sont des problèmes cliniques avec des questions aux étapes clés de la résolution de problème³⁵. Ils ont été validés pour l'évaluation du raisonnement clinique pour le diagnostic et la prise en charge d'un problème clinique^{35,36}. La revue de Nayer et al. synthétise les 12 conseils pour construire des KFP dans le but évaluer le raisonnement clinique³⁵. Cette méthode d'évaluation a été largement utilisée dans le domaine médical et est actuellement recommandée en France dans la nouvelle réforme des études médicales (R2C).

1.3.8. Les tests de concordance de script (TCS)

Le test de concordance de script (TCS) est bien validé pour explorer le raisonnement clinique en contexte d'incertitude³⁷⁻⁴⁰. Nous détaillerons ici son principe car il sera utilisé dans une des études menées pour cette thèse et décrite plus loin.

Il est basé sur la théorie des scripts qui dérive des théories cognitives d'apprentissage. Face à une situation clinique et dans un contexte de résolution de problème, un professionnel de santé utilise des scripts mentaux contenus dans sa mémoire à long terme⁴¹. Le TCS correspond à une courte vignette clinique qui constitue un problème car les données présentées sont insuffisantes ou soulèvent une ambiguïté (Figure 4). Puis on apporte à l'étudiant des informations complémentaires qui vont entraîner un micro-jugement de sa part en termes de diagnostic, d'investigation (biologie, radiologie...) ou de traitement. Le candidat doit alors choisir entre les différentes propositions formulées à l'aide d'une échelle de Likert entre -2 et + 2 (c-à-d de « beaucoup moins probable » à « beaucoup plus probable » ou encore de « la moins appropriée » à « la plus appropriée »). En pratique, le TCS comporte 3 éléments qui correspondent aux étapes de :

- la génération d'hypothèse (« Si vous pensiez à . . . » : option pertinente)
- la collecte des données (« . . . puis vous trouvez . . . » : nouvelle donnée)
- et l'interprétation des données (« . . . cette hypothèse devient . . . » avec une échelle de Likert).

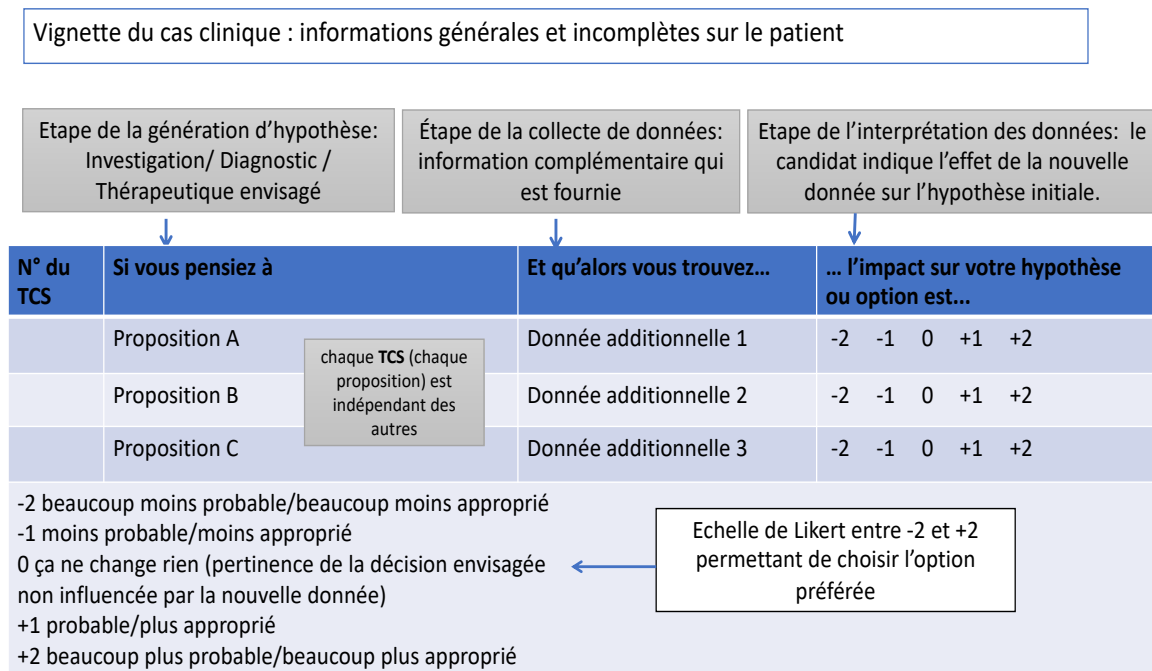


Figure 4 : Principes des tests de concordance de scripts

Les études publiées montrent qu'il faut environ une soixantaine de questions (20 vignettes de 3 questions chacune) pour atteindre une bonne fidélité de mesure (coefficient de validité interne au moins supérieur à 0,70)³⁷. Puis un panel d'experts, constitué de professionnels expérimentés dans le domaine envisagé (10 à 20) permet d'élaborer la grille de correction. Un groupe de relecture, indépendant du panel, peut être utile pour modifier certaines questions ou en supprimer d'autres, apparaissant non pertinentes d'après les réponses du panel (par exemple du fait de réponses trop univoques ou au contraire trop dispersées). Enfin, l'étudiant est évalué sur la base d'un score qui mesure le degré de concordance entre ses réponses et celles du panel d'experts. Le calcul du score pour un item du TCS se fait en fonction du nombre d'experts et de la distribution de leurs réponses par item. Par exemple, si parmi 10 experts, 8 d'entre eux ont répondu -2 et deux ont répondu -1 à un TCS donné, le score est alors $8/8 = 1$ à la réponse -2, $2/8 = 0,25$ à la réponse -1 et $0/8 = 0$ à la réponse 0, +1 et +2.

Ainsi, les TCS permettent d'évaluer le raisonnement clinique de l'étudiant dans un contexte d'incertitude. Il ne s'agit pas d'évaluer uniquement le diagnostic ou le traitement mais d'évaluer une étape au sein du processus du raisonnement puisque c'est un micro-jugement, à un moment donné du raisonnement, qui est évalué. Le TCS a une bonne validité de construit, c'est à dire qu'il existe une adéquation entre les scores obtenus et le degré d'expertise³⁹. L'utilisation des TCS peut donc se faire dans un cadre sanctionnant ou formatif. Il a également été validé dans la plupart des domaines médicaux mais également chez les infirmiers^{37 41,42}. Cette méthode

d'évaluation est actuellement recommandée en France dans la nouvelle réforme des études médicales (R2C).

1.3.9. « Instrument » d'autoévaluation du processus du raisonnement clinique

Des échelles psychométriques basées sur le modèle du raisonnement clinique ont été décrites par différentes équipes. Ces échelles sont des auto-évaluations du processus de raisonnement clinique par les apprenants en y reprenant les différentes étapes.

Koivisto et al.⁴³ ont créé une échelle psychométrique basée sur le modèle du raisonnement clinique de Levett-Jones et al.²⁰. L'échelle comporte 14 variables explorant les différentes étapes du raisonnement clinique en les cotant de 1 à 5 (Tableau 2). Cette échelle a été utilisée pour évaluer la perception d'apprentissage du raisonnement clinique chez les étudiants infirmiers lors d'un jeu sérieux. La validité interne de cette échelle était bonne avec un alpha de Cronbach à 0,647-0,832.

Liou et al.⁴⁴ ont également développé le même type d'échelle psychométrique (*Nurse clinical reasoning scale*) basée sur le modèle du raisonnement clinique décrit par Levett-Jones et al. Elle comporte 15 items cotés de 1 à 5.

Puis, Liaw et al.⁴⁵ ont évalué le raisonnement clinique dans le contexte de la détection de l'aggravation clinique d'un patient en développant une échelle (*CREST tool*) comportant 11 items qui combine ceux précédemment décrits et d'autres plus spécifiques à la détection de l'aggravation clinique.

Tableau 2 : Autoévaluation du processus d'apprentissage du raisonnement clinique (d'après Koivisto et al.⁴³, traduction française par A Blanié, D Benhamou)

J'ai appris à :		Entourez votre choix : 1 (pas du tout d'accord) à 5 (tout à fait d'accord)				
Collecter l'information	Collecter des informations en interrogeant le patient	1	2	3	4	5
	Collecter des informations en observant le patient	1	2	3	4	5
	Collecter des informations en mesurant des données du patient	1	2	3	4	5
Traiter l'information	Analyser les données pour parvenir à comprendre les signes et symptômes	1	2	3	4	5
	Distinguer l'information pertinente et non pertinente	1	2	3	4	5
Identifier les problèmes et solutions	Faire un diagnostic infirmier	1	2	3	4	5
	Prendre une décision de soins du patient indépendamment	1	2	3	4	5
	Prendre une décision de soins du patient en coopération avec les autres collègues	1	2	3	4	5

	Prendre une décision de soins du patient rapidement	1	2	3	4	5
Établir des objectifs	Prioriser les soins nécessaires pour le patient	1	2	3	4	5
	Fixer des objectifs	1	2	3	4	5
	Planifier des interventions infirmières	1	2	3	4	5
Prendre des dispositions	Faire des interventions infirmières en fonction des symptômes	1	2	3	4	5
Évaluer le devenir	Évaluer l'efficacité des interventions	1	2	3	4	5

Les preuves de validité des différentes échelles psychométriques évaluant la perception d'apprentissage du raisonnement chez les infirmiers sont présentées dans le tableau 3.

Tableau 3: Validité des échelles psychométriques évaluant la perception d'apprentissage du raisonnement chez les infirmiers (preuves de validité présentées selon le cadre théorique proposé par Messick⁴⁶)

	L'équipe de Koivisto et al. ⁴³	L'équipe de Liou et al. ⁴⁴ <i>Nurse clinical reasoning scale</i>	L'équipe de Liaw et al. ⁴⁵ <i>CREST tool</i>
Preuve du contenu	Validité de contenu	Validité de contenu	Validité de contenu
Relation avec les autres variables	Validité du construit réalisée mais non détaillée	Validité du construit du score entre infirmiers (n=100) 55,1±7,8 vs étudiants infirmiers (n=151) 52,6±7, p<0,05	Validité du construit du score entre groupe étudiants 2 ans (n=15) vs 3 ans (n=15) Validité de prédiction : comparaison des résultats entre l'échelle CREST vs l'échelle eRapid
Structure interne	Fiabilité interne : Cronbach's alpha= 0,647-0,832.	Fiabilité interne : Cronbach's alpha= 0,93 Fiabilité intraclasse (test et retest à 2 semaines) : (ICC 0,85)	Fiabilité interne : Cronbach's alpha= 0,92 (IC 0,86-0,95) Fiabilité interclasse : ICC =0,88

D'autres échelles psychométriques ont également été développées dans d'autres domaines comme dans le cadre des patients polytraumatisés⁴⁷.

Dans les études mise en place pour cette thèse, le questionnaire utilisé a été basé sur celui de Koivisto et al.²⁶ car les questions semblaient plus claires à comprendre pour les étudiants infirmiers. Il a été modifié en ajoutant des questions spécifiques sur le

problème de la détérioration du patient, l'étude de Liaw et al.²⁸ ayant été publiée ultérieurement.

1.3.10. Autres modèles d'évaluation du raisonnement clinique.

Le raisonnement clinique étant complexe, la recherche de méthodes d'évaluation est donc toujours en cours. Récemment, Cook et al. ont présenté un modèle d'évaluation clinique qui correspond à un arbre décisionnel à partir d'une situation clinique avec des mini tests à différentes étapes. Mais ce modèle, à notre connaissance, n'a pas encore été testé⁴⁸.

Charlin et al. ont utilisé un logiciel pouvant décrire une cartographie cognitive semi structurée avec différents types de connaissance et de liens (le modèle MOT plus software, LICEF (Laboratoire en Informatique Cognitive et Environnements de Formation, Research Center, Montreal, Quebec,) (Figure 5)⁴⁹. Ce logiciel ne semble pas utilisé en pratique courante.

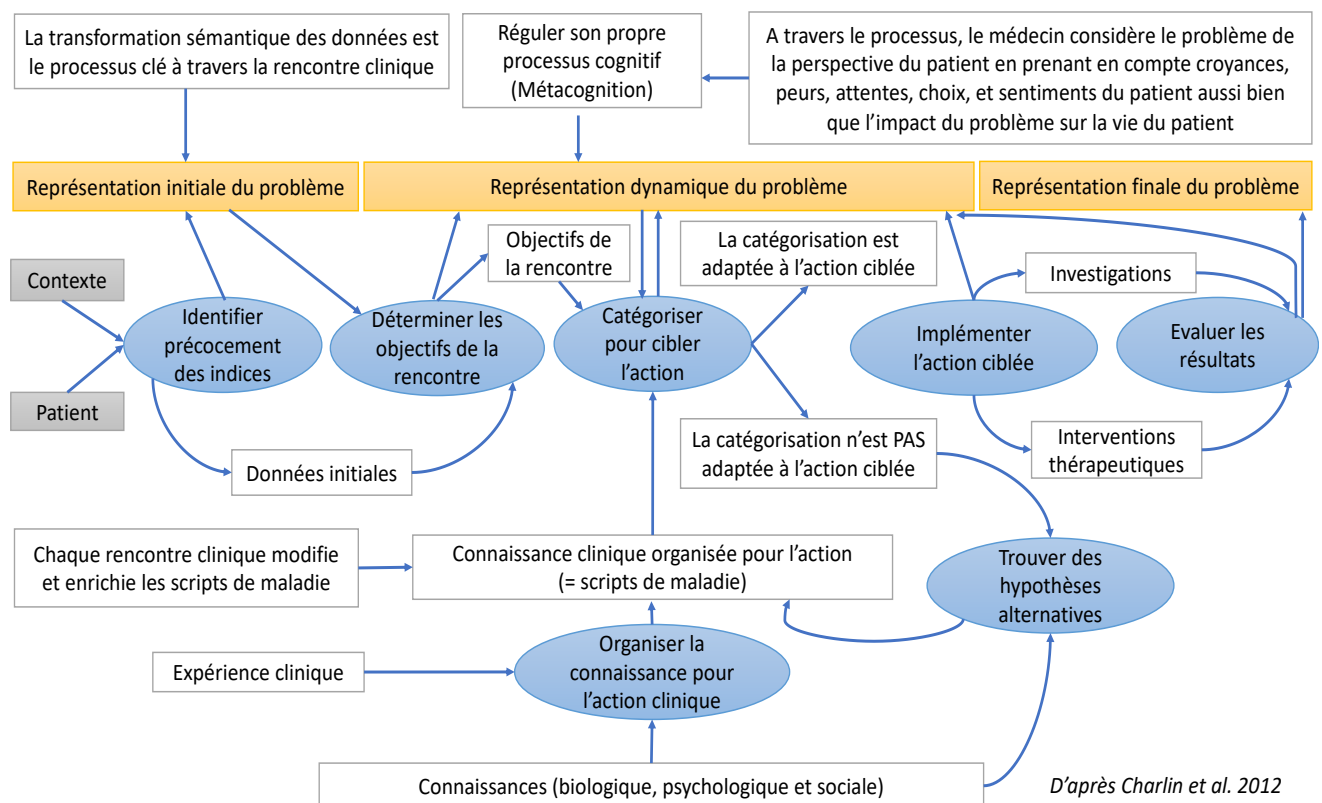


Figure 5 : Représentation graphique du processus du raisonnement clinique selon Charlin et al.⁴⁹.

Enfin, les développements technologiques, et plus particulièrement celui de l'intelligence artificielle, vont probablement bouleverser l'évaluation du raisonnement clinique dans les prochaines années⁵⁰⁻⁵².

En conclusion, le raisonnement clinique est complexe et multidimensionnel. Son évaluation implique l'utilisation d'instruments complémentaires.

2. Les jeux sérieux en santé

2.1 Définitions

Le jeu sérieux est un outil pédagogique de simulation. Selon le guide des bonnes pratiques en matière de simulation en santé⁵³, « Les jeux sérieux (de l'anglais *serious games*) sont des applications développées à partir des technologies avancées du jeu vidéo, faisant appel aux mêmes approches de design et savoir-faire que le jeu classique (3D temps réel, simulation d'objets, d'individus, d'environnements...) mais qui dépassent la seule dimension du divertissement. Ils combinent une intention sérieuse, de type pédagogique, informative, communicationnelle, ou d'entraînement avec des ressorts ludiques. Ils sont en quelque sorte une déclinaison utile du jeu vidéo au service des professionnels ».

Selon le dictionnaire américain de simulation en santé de 2020⁵⁴, le jeu sérieux (*serious game*) peut être défini comme :

- « Une compétition mentale jouée avec un ordinateur conformément à des règles spécifiques, qui utilise le divertissement pour poursuivre des objectifs de formation, d'éducation, de santé, de politique publique et de communication stratégique ».
- « Un jeu conçu principalement dans le but de résoudre un problème. Les jeux sérieux ont un but éducatif explicite et soigneusement pensé, et ne sont pas destinés à être joué principalement pour le divertissement. Les jeux sérieux sont des simulations d'événements réels ou de processus conçus pour résoudre un problème. »

La définition d'un jeu sérieux est parfois ambiguë et comporte des interfaces avec le e-learning, la notion de patient virtuel, ou la simulation informatique⁵⁵. Les différentes définitions issues du dictionnaire américain de simulation en santé (*2020 Healthcare Simulation Dictionary—Second Edition*) sont reprises dans le tableau et montrent à quel point certains termes sont proches⁵⁴ (Tableau 4).

Les jeux sérieux nécessitent des éléments de ludification (*gamification*) pour être qualifiés de jeu comme par exemple : une compétition, un score, des récompenses...⁵⁶. De plus, le type de jeu peut être très différent : jeu d'aventure, jeu de plateau, quizz, simulation...

Dans la littérature en santé, un jeu sérieux est souvent défini comme un jeu dont l'objectif est sérieux via un matériel numérique, informatique⁵⁶⁻⁵⁸. Dans cette thèse, nous étudierons donc les jeux sérieux comme des jeux vidéo à visée pédagogique, c'est à dire une simulation virtuelle sur un matériel numérique associant des éléments de jeu et un objectif pédagogique. Et nous ciblerons les jeux sérieux en santé pour la formation des professionnels de santé (et non l'éducation des patients).

Tableau 4 Définitions issues du dictionnaire américain de simulation en santé ⁵⁴

	Définitions issues du dictionnaire américain de simulation en santé 2020	Traduction française
Serious games Jeux sérieux	<i>A mental contest played with a computer in accordance with specific rules, which uses entertainment to further training, education, health, public policy, and strategic communication objectives (Zyda, 2005).</i> <i>A game designed for a primary purpose other than pure entertainment. Serious games have an explicit and carefully thought out educational purpose, and are not intended to be played primarily for amusement (Michael and Chen, 2006).</i> <i>Serious games are simulations of real-world events, or processes designed for the purpose of solving a problem.</i> <i>In the defense context, serious games are used to rehearse, train, or explore military options in a simulation of real-world events or processes (Australian Dept. of Defense).</i> <i>The "serious" adjective is generally appended to refer to products used by industries like defense, education, scientific exploration, healthcare, emergency management, city planning, engineering, religion, and politics.</i>	Une compétition mentale jouée avec un ordinateur conformément à des règles spécifiques, qui utilise le divertissement pour poursuivre des objectifs de formation, d'éducation, de santé, de politique publique et de communication stratégique Un jeu conçu principalement dans le but de résoudre un problème. Les jeux sérieux ont un but éducatif explicite et soigneusement pensé, et ne sont pas destinés à être joué principalement pour le divertissement. Les jeux sérieux sont des simulations d'événements réels ou de processus conçus pour résoudre un problème. Dans le contexte de la défense, les jeux sérieux sont utilisés pour répéter, entraîner et explorer les options militaires dans une simulation d'événements ou procédures du monde réel. L'adjectif « sérieux » est généralement ajouté pour se référer aux produits issus des industries telles que la défense, éducation, exploration scientifique, santé, gestion de l'urgence, planification de ville, ingénierie, religion et politique
Screen based simulation Simulation sur écran	<i>A simulation presented on a computer screen using graphical images and text, similar to popular gaming format, where the operator interacts with the interface using keyboard, mouse, joystick, or other input device.</i> <i>The programs can provide feedback to, and track actions of learners for assessment, eliminating the need for an instructor (Ventre & Schwid, in Levine Chapter 14).</i> <i>A computer-generated video game simulator that can create scenarios that require real-time decision making (Bonnetain, Biese, et al., 2009).</i>	Une simulation présentée sur un écran d'ordinateur utilisant les images et un texte, semblable au format de jeu de divertissement, dans lequel l'opérateur interagit avec l'interface en utilisant un clavier, une souris, un joystick ou un autre matériel pour réaliser les actions du jeu. Les programmes peuvent fournir un retour et traquer les actions des apprenants pour l'évaluation, supprimant la nécessité d'un instructeur. Un simulateur de jeu vidéo sur ordinateur qui peut créer des scénarios qui requièrent une prise de décision en temps réel.
Virtual patient Patient virtuel	<i>A representation of an actual patient. Virtual patients can take many forms such as software-based physiological simulators, simulated patients, physical manikins, and simulators, (Ellaway, Poulton, Fors et al., 2008).</i>	Une représentation d'un patient réel. Les patients virtuels peuvent prendre plusieurs formes telles que des simulateurs avec logiciel basé sur la physiologie, des patients simulés, des mannequins physiques et des simulateurs.

	<i>A computer program that simulates real-life clinical scenarios in which the learner acts as a health care provider obtaining a history and physical exam, and making diagnostic and therapeutic decisions (ASSH).</i>	Un programme informatique qui simule des scénarios cliniques véritables, dans lesquels un apprenant agit comme un prestataire de soins pour obtenir une histoire médicale, pratiquer un examen physique, émettre un diagnostic et prendre des décisions thérapeutiques.
Virtual reality simulation Simulation par réalité virtuelle	<i>Simulations that use a variety of immersive, highly visual, 3D characteristics to replicate real-life situations and/or health care procedures; virtual reality simulation is distinguished from computer-based simulation in that it generally incorporates physical or other interfaces such as a computer keyboard, a mouse, speech and voice recognition, motion sensors, or haptic devices</i>	Simulations qui utilisent une grande diversité de technologies immersive, hautement visuelles, avec des caractéristiques 3D afin de reproduire des situations de la vraie vie et/ou des procédures de soins ; la simulation à réalité virtuelle se distingue des simulations informatiques en ceci qu'elle inclut en général des interfaces physiques ou autres (comme un clavier d'ordinateur, une souris, la reconnaissance vocale, des capteurs de mouvements, du dispositifs haptiques) qui permettent de mieux reproduire l'action requise dans une situation ou contexte donnés.
Virtual simulation Simulation virtuelle	<i>The recreation of reality depicted on a computer screen (McGovern, 1994).</i> <i>A simulation involving real people operating simulated systems. Virtual simulations may include surgical simulators that are used for on-screen procedural training and are usually integrated with haptic device(s) (McGovern, 1994; Robles-De La Torre, 2011).</i> <i>A type of simulation that injects humans in a central role by exercising motor control skills (for example, flying an airplane), decision skills (committing fire control resources to action), or communication skills (as members of an air traffic control team) (Hancock et al, 2008).</i>	La reproduction de la réalité sur un écran l'ordinateur. Une simulation impliquant des personnes véritables qui opèrent des systèmes simulés. Les simulations virtuelles peuvent inclure les simulateurs chirurgicaux utilisés pour la formation procédurale à l'écran et généralement intégrés avec des instruments avec dispositif haptique); Un type de simulation qui implique un humain dans un rôle central à travers des compétences motrices (p.ex., piloter un avion), des compétences de prise de décision (p.ex. engager des ressources de contrôle du feu) ou des compétences de communication (p.ex., comme membres d'une équipe de contrôleurs aériens)
Computer-Based Simulation Simulation informatique	<i>The modeling of real-life processes with inputs and outputs exclusively confined to a computer, usually associated with a monitor and a keyboard or other simple assistive device (Textbook of Simulation). Subsets of computer-based simulation include virtual patients, virtual reality task trainers, and immersive virtual reality simulation (ibid)..</i>	Activités de simulation de procédure de la vie réelle qui sont effectuées au moyen exclusif d'un ordinateur avec un programme informatique, habituellement associé à un moniteur, et un clavier ou autre dispositif simple ; Les sous-ensembles de la simulation informatique incluent les patients virtuels, les simulateurs procéduraux en réalité virtuelle et la simulation par réalité virtuelle immersive.
Gamification	<i>The application of game design elements (conceptual building blocks integral to</i>	L'introduction d'éléments de jeu à un contexte non ludique.

Ludification	<i>building successful games) to traditionally non game contexts (Ruledge et al. 2018) The application of the characteristics and benefits of games to real world processes or problems. "gamification differs from serious games in terms of the design intention, with gamification interventions involving the application of game elements with a utilitarian purpose (Gentry et al. 2019)</i>	L'application des caractéristiques et des bénéfices des jeux aux processus et problèmes du monde réel. La ludification diffère des jeux sérieux en terme d'intention dans la conception, avec des interventions de ludification utilisant l'inclusion d'éléments de jeu avec un but utilitaire.
---------------------	--	---

Le développement des jeux sérieux est en plein essor. Malgré un coût encore élevé, les avantages de cet outil semblent multiples. D'une part, en termes de ressources requises car ils permettent une formation de masse avec un nombre limité de formateurs voire une formation à distance, et/ou grâce à l'utilisation de simples ordinateurs. D'autre part, ils permettent un apprentissage en toute sécurité pour le patient et l'apprenant. Cependant, les modalités et la valeur éducative des jeux sérieux dans la formation sont encore à préciser. Nous allons dans un premier temps rappeler quelques concepts d'apprentissage, puis faire une revue de la littérature de l'efficacité des jeux sérieux en santé puis dans le domaine du raisonnement clinique.

2.2 Jeux sérieux et théories d'apprentissage

En théorie, le jeu sérieux est un outil pédagogique intéressant car il associe les 4 piliers favorisant l'apprentissage habituellement décrits dans les sciences cognitives⁵⁹ : l'attention, l'apprentissage actif, le feedback et la consolidation.

Tout d'abord, l'attention est obtenue par un état d'alerte continu des apprenants pendant le jeu sérieux via le graphique, le son et/ou le défi posé par le jeu⁵⁹. La motivation du joueur est également essentielle pour maintenir cet état de vigilance pendant le jeu. Pour favoriser cette attention, on peut également guider l'apprenant, c'est-à-dire promouvoir des moyens qui vont l'aider à sélectionner les informations pertinentes (par des feedbacks ou autres stratégies)⁵⁹. Cependant, l'attention nécessite que l'apprenant reste concentré sur le jeu et il faut donc éviter que des éléments extérieurs ne viennent la perturber.

Ensuite, Il a bien été montré que la pédagogie active améliore efficacement la performance aux examens des étudiants dans la majorité des domaines de la formation scientifique⁶⁰. L'apprentissage actif est un processus d'apprentissage permis par une activité qui permet à l'apprenant de participer à l'enseignement et de ne pas rester passif. Or, un enseignement associant un jeu sérieux peut être assimilé à une pédagogie active puisque l'apprenant est, par définition, en interaction continue avec l'activité qu'est le jeu. Une méta-analyse sur l'efficacité de l'apprentissage des jeux de simulation hors santé a montré que les jeux dans lesquels les apprenants étaient activement engagés étaient plus efficaces que ceux où ils étaient passifs⁶¹.

Un autre pilier de l'apprentissage est le feedback⁵⁹. Dans les jeux sérieux, le feedback est présent et peut prendre différentes formes tels que le niveau de survie, un score,

ou des textes/feedbacks des actions du joueur au cours du jeu. Cependant, tous les feedbacks ne sont probablement pas équivalents en terme d'impact sur le processus d'apprentissage et certains pourraient même avoir des effets négatifs⁵⁹. La présence d'un débriefing avec un enseignant après la session de jeu sérieux est également un élément qui peut être réalisé mais a été peu étudié dans ce contexte⁶².

L'apprentissage peut également se consolider par la répétition⁵⁹. Grâce aux jeux sérieux, les apprenants ont la possibilité de s'entraîner à leur rythme, de répéter les cas, de voir d'autres variantes des cas. Tous ces éléments permettraient de renforcer l'ancrage mémoriel.

Concernant les autres théories d'apprentissage, les jeux sérieux permettent un apprentissage expérientiel décrit selon le cadre théorique proposé par Kolb ⁶³. Ce concept correspond à la succession d'une phase d'expérience concrète (lors du jeu) suivi d'une phase d'observation réflexive, puis d'un temps de conceptualisation (lors d'un débriefing). Dans les jeux sérieux qui correspondent à la prise en charge d'un patient virtuel dans un environnement également virtuel, l'étudiant va jouer son propre rôle. Il va acquérir de l'expérience, voir des scripts cliniques à travers les scénarios. Il peut voir des situations fréquentes mais aussi des situations rares, lui permettant de se projeter dans sa vie professionnelle future, de s'autonomiser dans un environnement sécurisé.

Selon la théorie de la pratique réfléchie d'Ericsson (*deliberate practice*), l'apprentissage nécessite un entraînement répétitif intentionnel mais aussi de promouvoir un feedback continu et itératif⁶⁴. Le jeu sérieux s'intègre totalement dans ce concept puisque l'apprenant peut répéter les parties de jeux pour corriger ses erreurs et peut recevoir ainsi un feedback autant de fois que nécessaire.

Ainsi, en théorie, le jeu sérieux semble un outil pédagogique idéal, mais quelles sont les preuves de son efficacité dans la littérature ?

2.3 Efficacité des jeux sérieux dans le processus d'apprentissage

Comme tout outil pédagogique, l'évaluation est nécessaire pour s'assurer de sa validité. Nous développerons plus en détail dans une autre partie de cette thèse, l'ensemble du processus de validation d'un outil pédagogique tel que le jeu sérieux. Nous allons pour le moment nous intéresser à un aspect de sa validation qui est la preuve de son efficacité dans l'apprentissage.

Un processus d'apprentissage peut être évalué selon les différents niveaux en suivant le modèle décrit par Kirkpatrick⁶⁵:

- La réaction (satisfaction) (niveau 1),
- L'acquisition de connaissances (niveau 2a) et compétences (niveau 2b)
- La rétention des connaissances/compétences sur une période de temps (niveau 2c)
- Le changement de comportement (impact professionnel) (niveau 3),
- et les résultats pour les patients (niveau 4)

En 2010, une première revue de la littérature de Cook et al. avait cherché à évaluer l'effet de l'emploi pédagogique des patients virtuels informatisés (*computerized-virtual patient*) en comparaison à aucune ou une autre intervention⁶⁶. Sur les 48 études incluses dans cette méta-analyse, 4 étaient qualitatives et 45 étaient quantitatives et comparatives (mais pas toujours randomisés). L'évaluation concernait les compétences (n=22 études), les connaissances (n=21), le raisonnement (n=15) et la satisfaction (n=12). En comparaison à l'absence d'intervention, l'emploi de la méthode du patient virtuel avait montré une amélioration forte de l'apprentissage. Sur les 5 études évaluant le raisonnement clinique en comparaison à l'absence d'intervention, l'effet était considéré comme fort. Il n'y avait pas de différence significative quand on le comparait à une autre intervention non informatisée. Cependant, ces résultats sont difficiles à interpréter car les études étaient très hétérogènes et anciennes (entre 1966 et 2008). De plus, la comparaison à l'absence d'intervention apporte peu en pédagogie. Enfin, cette méta-analyse était réalisée sur le patient virtuel informatisé ce qui ne correspond pas toujours à un jeu sérieux puisqu'il n'y a pas forcément d'élément de jeu. Depuis, une récente méta-analyse du groupe collaboratif d'éducation numérique à la santé ⁶⁷ a retrouvé 51 études randomisées contrôlées comparant la méthode du patient virtuel et à différentes méthodes pédagogiques. Cependant, seuls les patients virtuels sans haut niveau d'immersion (excluant les environnements virtuels 3D) étaient inclus. En comparaison à un enseignement traditionnel (n=25 études), les résultats sur l'apprentissage du patient virtuel en termes de connaissance n'étaient pas significativement différents mais une amélioration forte des compétences étaient retrouvée. Au sein de cette méta-analyse, 4 études concernaient le raisonnement clinique.

En 2015, la revue systématique de Wang et al. ⁵⁶, dans les suites de celle faite par Graafland et al.⁵⁸, avait ciblé les études, publiées entre 2007 et 2014, incluant des jeux sérieux dans la formation des professionnels de santé et la présence ou non d'évaluation. Sur les 42 jeux sérieux inclus, les thématiques abordées dans ces jeux étaient très variées avec principalement : la chirurgie (n=9 études), la pédiatrie (n=4), la médecine interne ou la médecine générale (n=3), les soins infirmiers (n=3), et la réanimation cardiopulmonaire (n=2). Le jeu pouvait être sous forme de simulation (n=16), de quizz seul (n=17), de puzzle (n=3) ou de différentes associations. Concernant l'évaluation, 79% (n=33) des études évaluaient le jeu sérieux comme un outil pédagogique. Cependant, toutes ces études étaient très hétérogènes : le type et le nombre de participants, le design (pré-post test, avec ou sans contrôle) et la méthode d'évaluation. Enfin, la qualité méthodologique des études était assez faible avec 10,5/18 en moyenne selon le score MERQSI (entre 6 à 16,5 sur 18).

Depuis, des revues de la littérature ont montré l'intérêt des jeux sérieux dans la formation des professionnels de santé^{57,68,69}. Ainsi, la revue systématique récente de Gentry et al. a évalué l'efficacité des jeux sérieux (*serious game/gamification*) dans la formation des professionnels de santé en comparaison à un enseignement traditionnel (19 études randomisées), à un autre enseignement digital (7 études randomisées) ou à un autre type de jeu (4 études randomisées)⁵⁷. Comme précédemment, compte tenu

de l'hétérogénéité des populations, des interventions, des comparaisons, du haut risque de biais de la majorité des études, une méta-analyse n'a pas pu être réalisée. Leur analyse de la littérature a cependant montré que le jeu sérieux apparaissait au moins aussi efficace que la méthode contrôle et, que dans la plupart des études, il était plus efficace pour améliorer la connaissance, les compétences et la satisfaction. En comparaison à l'enseignement traditionnel, le jeu sérieux améliorait la satisfaction mais la preuve était limitée (1 étude⁷⁰). En termes d'apprentissage des connaissances, un effet modéré à fort du jeu sérieux a été retrouvé dans 5 études en comparaison à un enseignement traditionnel, et un effet équivalent dans 2 études. En termes d'apprentissage des compétences, un effet faible à fort du jeu sérieux a été retrouvé dans 10 études en comparaison à un enseignement traditionnel, et un effet équivalent dans 4 études. Il y avait une preuve limitée de l'effet du jeu sérieux sur les attitudes des professionnels. Par contre, ces résultats étaient moins certains en comparant le jeu sérieux et un autre enseignement digital.

Depuis la revue de Gentry et al., d'autres études ont été publiées qui comparaient principalement les jeux sérieux à un autre enseignement digital. L'étude de Drummond et al. n'a retrouvé aucune différence significative entre le jeu sérieux et l'enseignement en ligne, en termes de prise en charge de l'arrêt cardiaque pour les étudiants en médecine ⁷¹. De même, l'étude de Dankbaar et al. a trouvé une amélioration des connaissances mais sans différence significative entre le jeu sérieux (incluant des vidéos) et un e-module (incluant des textes) chez les étudiants en médecine dans le domaine de la sécurité du patient⁷². Par contre, l'étude de De Sena et al. a retrouvé des meilleurs résultats théoriques et pratiques avec l'enseignement par des vidéos en ligne par rapport au jeu sérieux concernant la prise en charge de l'arrêt cardiaque ⁷³.

Une revue récente de la littérature de Foronda et al. sur la simulation virtuelle dans la formation des infirmiers, a trouvé 80 articles dont 69 (86%) montraient un bénéfice dans l'apprentissage des étudiants infirmiers⁷⁴. Cependant, le terme « *serious game* » ne faisait pas partie des mots clés du moteur de recherche qui utilisait uniquement le terme « *virtual simulation* ». La grande majorité des études (n=47, 59%) avaient évalué l'apprentissage des connaissances avec une amélioration avait été globalement retrouvée dans les études. En ce qui concerne la pensée critique (« *critical thinking* »), qui peut être considérée comme un équivalent de raisonnement clinique, seules 10 études (12,5%) avaient évalué son apprentissage. Cela correspondait principalement à la prise de décision (dans 7 études), au jugement clinique (n =1), aux tâches infirmières d'urgence et à la reconnaissance de la dégradation clinique d'un patient (n=1), à la reconnaissance de la pauvreté (n=1). Une autre méta-analyse récente du groupe collaboratif d'éducation numérique à la santé⁷⁵ a comparé la réalité virtuelle à différentes méthodes pédagogiques dans la formation des professionnels de santé. La réalité virtuelle était ici définie comme une technologie permettant d'immerger le joueur à différents niveaux grâce à des écrans d'ordinateur ou dans des salles de réalité virtuelle avec des casques. Donc seules quelques études correspondaient à notre définition des jeux sérieux. En comparaison à l'enseignement traditionnel, la réalité

virtuelle améliorait l'apprentissage des connaissances avec un effet modéré (8 études) et les compétences avec un effet fort (seulement 4 études).

Ainsi, les différentes terminologies rendent difficile une lecture simple de la littérature. Cependant, les jeux sérieux semblent efficaces comme outil pédagogique en comparaison à un enseignement traditionnel, mais pas nécessairement plus efficaces en comparaison à un autre enseignement numérique.

Le jeu sérieux a été étudié dans de nombreux domaines : pharmacie, chirurgie, réanimation cardio-respiratoire, pédiatrie, soins infirmiers... Certaines équipes se sont également intéressées à son utilisation pour améliorer les compétences non techniques. Une revue systématique récente sur les compétences non techniques dans le domaine de la simulation par réalité virtuelle (dont les jeux sérieux) a observé que les domaines étudiés étaient principalement la communication, le travail en équipe et la conscience de la situation ⁷⁶. Bien que le nombre d'études publiées augmente, les études évaluaient surtout l'usage et l'acceptabilité de la simulation par réalité virtuelle et peu d'études mesuraient l'effet de la réalité virtuelle sur l'apprentissage des compétences non techniques. Par exemple, Youngblood et al. ont montré une amélioration de la performance en termes de leadership (évaluation externe par l'échelle «*emergency medicine crisis resource management*»), sans différence significative entre le groupe Virtual VR et le groupe simulation haute-fidélité lors de la prise en charge de polytraumatisés aux urgences⁷⁷. Plus récemment, l'étude de Barré et al. a retrouvé une amélioration de la perception d'apprentissage du leadership et du travail en équipe en comparant les résultats obtenus avant et après une formation par simulation informatique répétée sur l'hémorragie du post-partum chez des sages-femmes⁷⁸. Cependant, cette étude n'avait pas de groupe contrôle et les sessions avec la simulation informatique étaient associées à un débriefing ciblant les compétences non techniques. De même, l'utilisation d'une simulation informatique sur l'hémorragie du post-partum qui ciblait les compétences non techniques des sages-femmes a permis d'améliorer ces compétences grâce à des scénarios de simulation haute-fidélité par rapport à une stratégie associant la même simulation informatique mais ne ciblant pas les compétences non techniques⁷⁹.

Concernant l'apprentissage du raisonnement clinique, le chapitre suivant présente une revue de la littérature des évaluations faites dans les jeux sérieux.

2.4. Le raisonnement clinique dans les jeux sérieux

Le jeu sérieux semble être un outil pédagogique de choix pour l'apprentissage du raisonnement clinique. En effet, les scénarios virtuels permettent l'apprentissage expérientiel décrit par Kolb ⁶³. Il y a alors une acquisition de scripts cliniques (groupement d'informations générales éventuellement incluses au sein d'une séquence d'évènements et relatives à une situation médicale donnée) et de démarches de raisonnement.

L'équipe de Posel et al. a décrit 12 conseils pour mettre en avant le potentiel de la méthode du patient virtuel dans le développement des compétences en raisonnement clinique⁸⁰. L'emploi d'un patient virtuel permet de : (1) créer précisément le cas dont on a besoin du fait de la flexibilité du virtuel ; (2) mettre en avant l'importance de la preuve dans la prise de décision ; (3) mettre en avant l'acquisition des données du patient ; (4) mettre en avant l'importance de faire des exposés récapitulatifs du cas ; (5) permettre de développer le champ lexical médical en introduisant différents éléments de sémantique ; (6) cibler les diagnostics différentiels ; (7) souligner l'importance de la communication ; (8) inclure un feedback continu et immédiat ; (9) apprendre de ses erreurs dans un environnement sécurisé ; (10) utiliser le patient virtuel dans différentes modalités pédagogiques ; (11) faire des apprenants des auteurs de cas de patients virtuels comme un exercice ; (12) faire une étape de réflexion après les cas comme lors d'un débriefing. La majeure partie de ces conseils pourrait être repris pour les jeux sérieux.

Depuis plusieurs années, des équipes ont développé des jeux sérieux en affichant le raisonnement clinique ou certaines étapes (par exemple : la prise de décision) comme objectifs pédagogiques de leur jeu, mais souvent sans évaluer réellement l'apprentissage de ce raisonnement clinique. Par exemple, l'équipe norvégienne de Johnsen et al. a développé un jeu sérieux pour développer la formation au raisonnement clinique devant une décompensation de bronchite chronique chez les étudiants infirmiers. Cependant, à notre connaissance, seule des études d'usage du jeu et de la perception de son intérêt ont à l'heure actuelle été réalisées^{81,82}.

Une revue de la littérature des études évaluant le raisonnement clinique dans le cadre des jeux sérieux chez les professionnels de santé a été réalisée pour cette thèse et est présentée dans le tableau 5. La définition de l'évaluation du raisonnement clinique a été difficile à établir. Elle comprenait soit une évaluation du résultat (bon diagnostic, traitement, tri) soit une évaluation plus globale qui peut prendre la forme d'une échelle de compétence (par exemple de la détection de l'aggravation d'un patient [avec quelques items évaluant le raisonnement]), d'une autoévaluation du processus du raisonnement clinique ou encore de tests de concordance de scripts. Les études de validité des jeux sérieux qui n'évaluaient pas l'efficacité en termes d'apprentissage du raisonnement clinique n'ont pas été incluses.

Entre 2008 et février 2020, nous avons retrouvé 14 études randomisées contrôlées, 3 études comparatives non randomisées et 5 études de cohorte ou avant/après. Certaines équipes ont évalué plutôt le résultat avec la prise de décision (n=6 études). Par exemple, l'étude allemande de Chon et al. présentait son jeu sérieux « *Emerge* » comme un outil développant le raisonnement clinique des étudiants en médecine devant un patient présentant un tableau d'abdomen aigu aux urgences. Les méthodes d'évaluation de son efficacité étaient faites par des QCM de connaissances et des questions explorant les diagnostics et les traitements à partir de cas texte qui provenaient d'ECOS¹⁶. Ils n'ont retrouvé qu'une amélioration des connaissances mais pas de la capacité diagnostique, et il n'y avait pas de groupe contrôle. Dans le jeu de cas virtuel de consultation sur le cancer de l'œsophage (*ALICE*) pour les étudiants en

médecine, Kleinert et al. évaluaient si le bon diagnostic et le bon traitement étaient donnés. Ils retrouvaient une amélioration après l'utilisation du jeu (par rapport à l'évaluation avant son emploi), mais l'étude n'avait pas non plus de groupe contrôle⁸³. Les études américaines de Mohan et al.⁸⁴ et de Knight et al.⁸⁵ se sont intéressées à l'entraînement à la décision au tri chez les patients traumatisés à leur arrivée aux urgences. Une meilleure décision de triage de trauma a été retrouvée dans ces 2 études avec le jeu sérieux par rapport au groupe contrôle.

D'autres équipes ont utilisé des méthodes d'évaluation plus globale du raisonnement (n=16 études). Bogossian et al. ont développé un programme de formation à la détection de l'aggravation clinique d'un patient pour les étudiants infirmiers (*First2ACT WEB*). Cependant leur étude évaluait les connaissances et la perception d'apprentissage de connaissances et de compétences à la détection de l'aggravation ce qui ne correspond pas nécessairement à une évaluation du raisonnement clinique⁸⁶. Dans le cadre de l'apprentissage du processus du raisonnement clinique, l'équipe de Koivisto et al. a utilisé une grille d'autoévaluation détaillée de ce processus de raisonnement⁴³. Après utilisation du jeu sérieux ciblant la détection de l'aggravation d'un patient en postopératoire, une bonne perception de cet apprentissage chez les infirmiers a été observée. Mais dans les 2 études précédentes, aucune n'avait un groupe contrôle et l'étude de Koivisto et al. n'avait pas de pré-test^{43,86}. Liaw et al. ont utilisé une échelle RAPIDS comprenant une partie sur le domaine de l'approche ABCDE et une autre sur la communication SAED⁸⁷. L'approche ABCDE est une approche standardisée pour prioriser la prise en charge des patients dans un état critique en utilisant le moyen mnémotechnique suivant : A *Airway*, B *Breathing*, C *Circulation*, D *Disability* and E *Exposure*. Une amélioration du score « après » par rapport à « avant » a été retrouvée mais sans différence significative par rapport à la simulation haute-fidélité. Seule une étude a utilisé les tests de concordance de script comme évaluation du raisonnement clinique en comparant deux modalités pédagogiques différentes de jeu sérieux pour la formation des étudiants à la prise en charge d'urgences abdominales : des cas nombreux et courts (éléments clés) versus des cas peu nombreux et longs (analyse systématique)⁸⁸. Aucune différence n'a été retrouvée entre ces deux modalités pédagogiques.

Ainsi, dans la littérature, les études sur le raisonnement clinique dans le cadre des jeux sérieux restent hétérogènes et limitées en nombre. Globalement, il semble y avoir une amélioration du « raisonnement clinique » avec le jeu sérieux mais cet avantage est encore incertain quand on le compare à d'autres méthodes pédagogiques.

Tableau 5 : Revue de la littérature des études évaluant le raisonnement clinique dans le cadre des jeux sérieux chez les professionnels de santé (RC= Raisonnement clinique ; JS= Jeux sérieux, RCT : étude randomisée contrôlée).

Études	Type d'étude	Population (n inclus)	Thème	Type d'intervention	Contrôle	Durée-fréquence intervention	Méthode d'évaluation du RC	Résultats concernant le RC	Commentaires/Limites
Knight et al. 2010⁸⁵	RCT	Professionnels de santé (n=91)	Prise en charge des polytraumatisés pour l'entraînement au tri	JS " <i>Triage trainer</i> "	Carte d'exercice	60min - 1 fois	(1) Tri correct par rapport à la situation (2) temps mis pour le tri	Meilleur tri correct avec JS par rapport au contrôle	
LeFlore et al. 2012⁸⁹	RCT	Étudiants Infirmiers (n=93)	Prise en charge d'une détresse respiratoire en pédiatrie	JS " <i>Virtual Patient Trainer</i> "	Lecture	2-3h - 1 fois	(1) QCM des connaissances (2) temps pour appliquer les connaissances lors de 2 ECOS	Meilleure acquisition des connaissances et plus de rapidité pour la mise en application avec JS	Différence de temps faible (1min), pas réellement du RC
Liaw et al. 2014⁸⁷	RCT	Étudiants infirmiers (n=57)	Détection de l'aggravation clinique d'un patient	JS " <i>eRapids</i> "	Mannequin Haute-fidélité	2h-1 fois	Immédiatement après et à 2,5 mois : échelle RAPIDS comprenant une partie sur domaines ABCDE et SAED	Amélioration de la détection de l'aggravation du patient avant/après (immédiatement après et à 2,5 mois) mais non significatif entre JS et Haute-fidélité	Évaluation de la rétention (car les 2 groupes avaient eu une simulation Haute-fidélité 8 mois avant)
McGrath et al. 2015⁹⁰	RCT	Internes de médecine d'urgence (n=35)	Prise en charge des patients aux urgences	Examen par un examinateur en simulation virtuelle	Examen avec un oral traditionnel en face à face par un examinateur	?	Grille basée sur l' <i>american board of emergency medicine</i> évaluant 10 actions critiques et 8 compétences (dont l'acquisition de données, résolution de problème...)	Aucune différence de score observée entre les 2 groupes montrant la faisabilité d'utiliser une évaluation virtuelle (mais l'examinateur jouait l'avatar)	Évaluation sanctionnante (test un examen)
Bogossian et al. 2015⁸⁶	Avant/ Après	Étudiants infirmiers (n= 367)	Détection de l'aggravation clinique d'un patient	Programme " <i>First2ACT WEB</i> "	Non	?	(1) score de performance du jeu (2) QCM des connaissances (3) autoévaluation de la perception des connaissances	Amélioration des connaissances et de l'autoévaluation des connaissances et	Validation du score en faisant l'association score de performance et connaissance.

							et des compétences, et confiance dans la détection de l'aggravation d'un patient	compétences, et de la confiance dans la détection	Pas de groupe contrôle.
Kleinert et al. 2015²⁹	Avant/ Après	Étudiants en médecine (n=62)	Consultation pour cancer de l'œsophage	JS "ALICE"	Non	?	(1) parcours, (2) diagnostic, (3) traitement : corrects/incorrects	Amélioration avant/après du diagnostic et traitement corrects	Pas de groupe contrôle
Lin et al. 2015⁹¹	Cohorte	Novices, internes, seniors de chirurgie (n= 49)	Prise de décision d'urgences chirurgicales	JS "SICKO"	Non	1 session	Autoévaluation des experts sur l'intérêt du JS dans plusieurs étapes du RC	Bonne perception de l'intérêt du JS pour l'apprentissage des compétences cliniques dont certaines étapes du RC	Mais perception. Étude évaluant la validité du JS (construit du score validé)
Koivisto et al. 2016⁴³	Cohorte	Étudiants infirmiers (n= 166)	Détection de l'aggravation clinique d'un patient en postopératoire	JS	Non	30-40min	Autoévaluation par une grille détaillée du processus de RC	Bon score en terme de perception d'apprentissage du processus de RC	Pas de groupe contrôle et pas de pretest
Dankbaar et al. 2016⁹²	RCT	Étudiants en médecine (n=61)	Prise en charge d'urgences aux urgences	Groupe jeu : e-module + JS "AbcdeSIMS"	(2) Groupe e-module + cas texte sur ordinateur (3) Groupe contrôle : e module seul	e-module + 2h pour JS ou cas texte	Évaluation sur patient simulé : (1) échelle de compétences cliniques (Items de l'approche ABCDE et diagnostic), (2) échelle de compétences en communication (3) échelle de performance globale	Bon score de compétences (clinique/communication/globale) mais sans différence significative entre les 3 groupes (mais tous ont reçu une formation par e-module)	Évaluation motivation et perception de la charge cognitive
Liaw et al. 2017⁹³	RCT	Infirmiers (n=64)	Détection de l'aggravation clinique d'un patient	Web based educationnal program (e-Rapids) : video + documents sur ABCDE et SAED +	Contrôle : pas de formation	2,5-3h	1 semaine après : (1) QCM des connaissances (2) Échelle RAPIDS évaluant détection et prise en charge de la détérioration ; et transmission de la détérioration sur des scénarios de simulation	Amélioration dans le groupe intervention des connaissances, de la performance à la détection/prise en charge d'une détérioration et également à sa transmission ainsi qu'une meilleure	Amélioration minime significative dans contrôle de la performance en détection/ prise en charge probablement dû à l'effet pretest sur

				simulation patient virtuel			(3) % de mesures des signes vitaux	proportion de mesures des signes vitaux entre avant/après la formation web.	mannequin. Comparaison à aucune autre formation
Schubach et al. 2017⁸⁸	Quasi RCT	Étudiants en médecine (n=56)	Prise en charge d'urgences abdominales	Groupe éléments clés : Patient virtuel avec de nombreux cas courts et instructions de cibler les éléments clés	Groupe systématique : Patient virtuel avec quelques cas longs et instruction d'être complet et systématique	2 sessions de 50min ; à une semaine d'intervalle	Tests de concordance de scripts (TCS)	Pas de différence significative de score aux TCS entre les groupes éléments clés et systématique (respectivement, 57,4 (50,9-63,9) vs 62,7 (57,2-68,2, p=0,214)	Présence d'un même temps de discussion avec un enseignant après chaque session sur les cas a pu annuler l'effet.
Graafland et al. 2017⁹⁴	RCT	Internes de chirurgie (n= 24)	Conscience de la situation en chirurgie	Formation traditionnelle + JS " <i>Dr Game, surgeon trouble</i> "	Formation traditionnelle (lecture et entraînement pratique sur cochon)	2j pour formation traditionnelle + 30 min X 2 pour JS	(1) évaluation sur cochon des compétences à reconnaître et résoudre des problèmes (2) temps pris pour la reconnaissance et résolution	Meilleure proportion de résolution de problème mais pas de différences significatives pour la reconnaissance et le temps	Faible nombre d'inclusion
Diehl et al. 2017⁹⁵	RCT	Médecins généralistes (n= 134)	Prise en charge de l'insulinothérapie chez les patients diabétiques	JS « <i>Insuonline</i> »	Formation continue traditionnelle	4h	Échelle de compétence comportant des QCM de connaissances et des cas cliniques (avec QCM sur des résolutions de problèmes, attitudes) avant / après et à 3 mois	Amélioration du score de compétence entre avant et après et entre avant et à 3 mois. Meilleure amélioration dans le groupe JS vs contrôle (+ 40% vs + 34%, p<0,01)	
Mohan et al. 2017⁸⁴	RCT	Médecins urgentistes (n= 368)	Prise en charge des polytraumatisés pour l'entraînement au tri	JS ' <i>Night shift</i> '	Programme standard		Prise de décision concernant le tri	Amélioration de la prise de décision de tri dans une simulation virtuelle.	

Dankbaar et al. 2017⁷²	Comparative (non randomisée)	Internes (n=159)	Prise en charge de patients aux urgences	Lecture + JS <i>AbcdeSIMS</i> (6 cas)	Lecture seule	10H + 2,5h si JS	Évaluation sur patient simulé avec (1) échelle de compétences cliniques (Items approche ABCDE et diagnostic), (2) échelle de compétence en communication (3) échelle de performance globale	Meilleur score de compétence clinique dans groupe JS vs lecture seule mais sans différence pour la performance globale ni compétences en communication ni motivation. Pas de différence à 2 semaines.	Non randomisée
Middecke et al. 2018¹⁴	Comparative (non randomisée)	Étudiants en médecine (n= 112)	Prise en charge de patients aux urgences	JS " <i>EMERGE</i> "	Groupe PBL : Apprentissage par problème (<i>problem-based-learning</i>)	90 min - 10 fois	(1) Questions clés de procédures diagnostiques et thérapeutiques sur 6 cas cliniques (2) score analysant les différentes procédures correctes ou non (diagnostic, traitement, examen clinique, interrogatoire, ordre des examens biologiques, transfert du patient) lors d'une session <i>Emerge</i> sur 4 cas	Meilleur score global sur les 6 cas du groupe JS vs PBL (62,5% vs 54,5%, p=P,0015). Meilleur score pour chacun des 4 cas de <i>Emerge</i> dans le groupe JS vs groupe PBL (mais le groupe JS savait déjà jouer avec)	Non randomisée (78 dans le groupe <i>Emerge</i> et 34 PBL) et plus d'exposition à des cas dans le groupe <i>Emerge</i> (40 cas) vs PBL (5 cas)
Mohan et al. 2018⁹⁶	RCT	Médecins urgentistes (n= 320)	Prise en charge des polytraumatisés pour l'entraînement au tri	Groupe JS format aventure " <i>Night shift</i> " (engagement narratif) et Groupe JS format puzzle " <i>shift: the next generation</i> " (raisonnement analogique)	Groupe contrôle actif (documents en texte) et Groupe contrôle passif (pas d'intervention)	2h-1 fois	Performance lors d'une simulation virtuelle (JS 'Night shift') mesurée par la proportion de sous triage (cas sévère non transféré dans centre de trauma comme recommandé).	Diminution du sous triage quelques soit le jeu par rapport au groupe sans intervention (-18% (aventure) et -17% (puzzle)). Pas de différence significative entre groupe document texte et pas d'intervention	Évaluation de la performance sur simulateur virtuel et non en réalité

Chung et al. 2018⁹⁷	Comparative (non randomisée)	Infirmiers (n=130)	Détection de l'aggravation clinique d'un patient	Programme "First2ACT WEB" (WB)	Simulation avec patient standardisé (F2F)	1h-1,5h (WB) et 2h (F2F)	(1) questionnaire de connaissances (2) performance sur 3 cas	Amélioration des connaissances dans les 2 groupes. Amélioration de la performance dans le groupe WB mais pas groupe F2F.	Cependant meilleure évaluation de la qualité du feedback dans groupe F2F
Haerling et al. 2018⁹⁸	RCT	Étudiants infirmiers (n=81)	Prise en charge d'une décompensation de BPCO	Simulation virtuelle "vSim"	Simulation Haute fidélité		(1) Questionnaire de connaissance (2) évaluation satisfaction et confiance en soi avant/après (n=81) (3) évaluation de la performance sur patient standardisé chez 25 étudiants dans 3 domaines : cognitif, affectif et psychomoteur grâce aux scores utilisant le <i>Lasater Clinical judgment Rubric</i> et le <i>Creighton Simulation evaluation instrument</i> (4) coût	Amélioration des connaissances et de la confiance en soi entre avant et après mais sans différence significative entre les groupes. Pas de différence significative des scores de performance entre les groupes. Coût moins important avec la simulation virtuelle en comparaison avec simulation haute-fidélité	
Chon et al. 2019¹⁶	Avant /Après	Étudiants en médecine (n= 150)	Prise en charge d'urgences abdominales	JS "EMERGE"	0		(1) QCM de connaissances (2) questions de connaissance de procédures diagnostiques et thérapeutiques sur un cas texte dérivé d'ECOS	Amélioration du QCM de connaissance mais sans amélioration des diagnostics corrects	Pas réellement RC

Padhila et al. 2019	RCT	Étudiants infirmiers (n=42)	Prise en charge détresse respiratoire	Simulation avec patient virtuel	Simulation avec cas basse fidélité	45min	Immédiatement après et à 2 mois: (1) satisfaction (2) QCM de connaissance ciblant RC (3) perception d'auto-efficacité	Meilleure satisfaction et amélioration des connaissances supérieure dans le groupe patient virtuel par rapport au groupe contrôle immédiatement après et à 2 mois. Pas de différence significative en terme de perception d'auto efficacité entre les groupes	Utilisation de QCM de connaissance pour évaluer RC
Midecke et al. 2020¹⁵	RCT	Étudiants en médecine (n=69)	Prise en charge de patients aux urgences	Groupe A= tronc commun de cas (>46 cas) par JS "EMERGE" + 2 cas différents	Groupe B= même tronc commun de cas par "EMERGE" + 2 autres cas différents Groupe AB= même tronc commun de cas par "EMERGE" + les 4 cas précédents	10 X 90min	Évaluation sur 4 cas du jeu du RC grâce à un score de jeu comportant différents items (histoire, examen clinique, examens complémentaire, traitement mis en place, diagnostic, transfert du patient...).	Pas de différence significative entre les groupes donc selon les auteurs la formation au RC via le JS n'était pas spécifique d'un cas mais pouvait être transférable à d'autres cas.	Limites : absence de validation préalable de la construction de leur score via le JS donc pas certain que leur score reflète le niveau de performance.

II - Contribution expérimentale

1. Problématique et hypothèses

Le soin clinique de qualité nécessite une succession d'analyses et de décisions qui constituent le raisonnement clinique. L'incidence des erreurs de diagnostic varie selon les spécialités médicales³⁻⁵. Parmi les causes des erreurs diagnostiques, il est apparu qu'environ 75 % des erreurs diagnostiques sont liées à un défaut du processus cognitif du clinicien⁸. Les difficultés de raisonnement clinique affectent entre 5 et 15 % des étudiants en médecine et il est souvent détecté tardivement⁹. Le raisonnement clinique, au cœur de la pratique médicale, est ainsi un véritable défi pédagogique. Il s'agit donc bien d'un mécanisme essentiel qu'il faut non seulement comprendre mais sur lequel il semble nécessaire d'agir^{10,11}. Des stratégies de remédiation ont été proposées par Audétat et al. pour chaque difficulté spécifique du processus de raisonnement clinique¹². Bien que certaines études aient tenté des actions correctrices, les approches cognitives sont encore rares, peu développées et les méthodologies encore imparfaites^{11,13}. Parmi celles-ci, les interventions éducatives, c'est-à-dire qui ont pour but de modifier en profondeur le raisonnement clinique, sont peu développées et leurs résultats inconstants. De plus, les méthodes de pédagogie innovante n'ont été que peu utilisées et c'est notamment le cas de la simulation et de l'emploi des méthodologies numériques¹⁴⁻¹⁶. Les jeux sérieux sont efficaces et peuvent, entre autres, cibler certaines compétences dont le raisonnement clinique⁵⁷. Cependant, la plupart des études sur le raisonnement clinique dans le cadre des jeux sérieux, restent subjectives avec des évaluations qualitatives ou des autoévaluations des apprenants ou concernent uniquement le résultat (prise de décision). Par conséquent, la valeur éducative et les modalités des jeux sérieux dans la formation du raisonnement clinique des professionnels de santé restent à approfondir.

Cette thèse avait pour objectif de tester un mode d'apprentissage utilisant la simulation par les jeux sérieux afin d'améliorer le raisonnement clinique chez les professionnels de santé.

Dans un premier temps, le jeu sérieux *LabForGames Warning* a été développé dans le centre de simulation LabForSIMS comme outil pédagogique pour les étudiants infirmiers en ciblant la détection de l'aggravation d'un patient et la communication.

Une première hypothèse était que ce jeu sérieux *LabForGames Warning* dédié à la détection de la dégradation d'un patient et la communication interprofessionnelle était un outil valide pour la formation des professionnels de santé. L'objectif de la première étude était d'évaluer les preuves de validité du jeu sérieux *LabForGames Warning*.

Une seconde hypothèse était que la simulation numérique utilisant le jeu sérieux *LabForGames Warning* était plus efficace qu'un enseignement traditionnel sur l'apprentissage du raisonnement clinique chez les étudiants infirmiers. L'objectif de la 2^{ème} étude a été de comparer deux modalités d'enseignement en comparant un

groupe formé par simulation numérique avec *LabForGames Warning* et un groupe formé par enseignement traditionnel sur l'apprentissage du raisonnement clinique

2. Développement du jeu sérieux *LabForGames Warning*

Le premier projet de développement de jeux sérieux, que j'ai coordonné, a été mis en place par les équipes du centre de simulation LabForSIMS de la Faculté de Médecine l'Université Paris-Sud/Saclay entre 2013 et 2015. Le jeu *LabForGames Neuro* qui ciblait les étudiants en médecine avait pour objectif pédagogique la prise en charge d'urgences neurologiques aux urgences.

Dans un second temps, le projet de jeu sérieux *LabForGames Warning*, sujet de la thèse, a été monté par l'équipe de LabForSIMS en partenariat avec les équipes des Instituts de formation en soins infirmiers (IFSI) du Sud Francilien, Perray Vacluse, Paul Guiraud et Etampes. J'ai coordonné l'ensemble du projet depuis sa phase initiale en 2014 jusqu'à sa finalisation en avril 2017. Le développement a été permis par l'obtention d'un financement via un appel à projet lancé par l'Agence Régionale de Santé (ARS) île de France.

La détection de l'aggravation d'un patient et la communication interprofessionnelle, enjeux majeurs de santé publique, ont été choisis comme objectifs pédagogiques du jeu sérieux. Le rôle central des infirmiers dans le système de soins implique qu'ils doivent être en mesure de reconnaître la détérioration clinique d'un patient mais également de donner l'alerte à une équipe médicale ce qui nécessite une bonne communication interprofessionnelle⁹⁹⁻¹⁰³. La méthode SBAR, traduite en français sous le terme SAED (Situation, Antécédents, Évaluation, Demande)¹⁰⁴, a été choisie comme outil de standardisation de cette communication entre professionnels de santé dans le jeu. La cible initiale de ce jeu étant l'étudiant infirmier, les objectifs pédagogiques devaient également être en accord avec les compétences et unités d'enseignement du référentiel de formation infirmier¹⁰⁵.

Le modèle virtuel de notre jeu est de type *Sims®* avec un avatar infirmier qui doit surveiller un patient dans différents contextes de soins, dans un environnement 3D (Figures 6). Le jeu a été nommé « *LabForGames Warning* » en référence au système d'alerte de détection précoce qui existe dans la littérature¹⁰⁰. La réalisation des cas a consisté en un processus de dialogue itératif entre l'équipe pédagogique et l'équipe de développement (Interaction Healthcare®). L'équipe pédagogique pluridisciplinaire était constituée d'experts formateurs en soins infirmiers des IFSI impliqués (IFSI Sud Francilien, Perray Vacluse, Paul Guiraud et d'Etampes et d'experts médicaux du CHU de Bicêtre (anesthésistes-réanimateurs). Les 4 scénarios élaborés ont été :

- Cas 1 : patient adulte hospitalisé en postopératoire après une prothèse totale de hanche programmée, se compliquant d'une hémorragie.
- Cas 2 : enfant de 1 an aux urgences pour une gastroentérite compliquée d'hypovolémie.
- Cas 3 : patient âgé en Établissement d'Hébergement pour Personnes Âgées Dépendantes (EHPAD) dément et sous anticoagulant, présentant une

dégradation neurologique progressive à 48h d'une chute mécanique avec traumatisme crânien.

- Cas 4 : patient schizophrène hospitalisé en psychiatrie, avec apparition d'une complication sous la forme d'un syndrome occlusif.

Le jeu n'était pas joué en temps réel mais une horloge permettait aux apprenants de suivre le déroulé temporel du cas. Pour chaque scénario, trois étapes consécutives cliniques (quasi normale, aggravation modérée puis sévère) étaient construites pour reproduire l'aggravation clinique d'une complication, avec pour objectif d'introduire le concept de signes d'alerte précoces¹⁰⁰.

Avant de jouer, l'apprenant suivait un tutoriel obligatoire pour la prise en main du logiciel. Les actions des étudiants étaient obtenues en cliquant sur différentes zones interactives. Par exemple, concernant l'examen clinique, l'étudiant pouvait prendre la pression artérielle, faire un examen de la motricité des membres... L'interrogatoire du patient était permis par un menu déroulant avec des questions pré-déterminées (certaines pertinentes, d'autres non) qui entraînaient des réponses verbales pré-enregistrées du patient. L'étudiant pouvait aussi avoir accès au dossier médical et infirmier (prescription, transmissions ciblées...). Il pouvait à tout moment appeler le médecin avec un téléphone et présenter le cas en lui suggérant d'utiliser la méthode SAED.

A la fin de chaque scénario, un feedback automatique était présenté à l'étudiant incluant les objectifs pédagogiques, les messages importants du cas, les référentiels et un score. Le score était à la fois global sur le cas et aussi détaillé selon une grille établie par l'équipe pédagogique lors de la phase de développement. Des actions concernant l'examen clinique (évaluation de la douleur, prise de pression artérielle...) et des décisions prises par l'étudiant (par exemple : appeler le médecin) étaient notées en points positifs, négatifs ou neutres en fonction des différentes étapes du jeu. De plus, des points positifs ou négatifs ont été attribués en fonction de la qualité de la communication pendant la partie de communication utilisant le SAED.

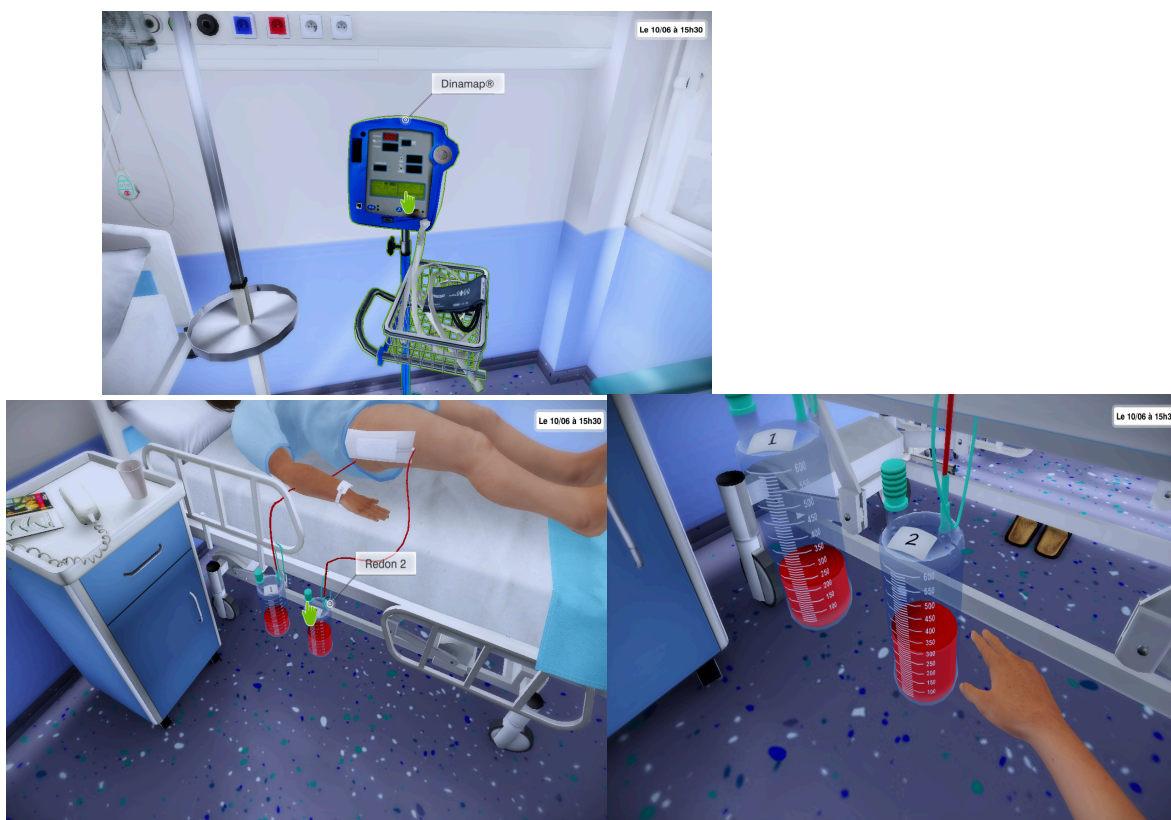
Sur le plan pratique, le jeu pouvait être joué à partir d'un ordinateur hors connexion (en téléchargeant l'application) ou en ligne (connexion internet) en accédant à une plateforme.

Figure 6 : Capture d'écran du jeu sérieux *LabForGames Warning*

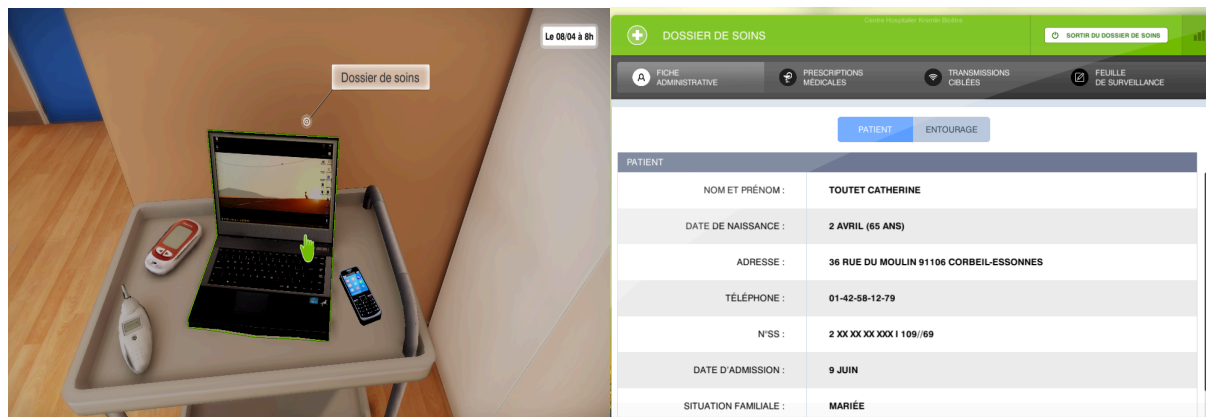
- a. Environnement du cas hémorragie post-opératoire (gauche) et trauma crânien (droite)



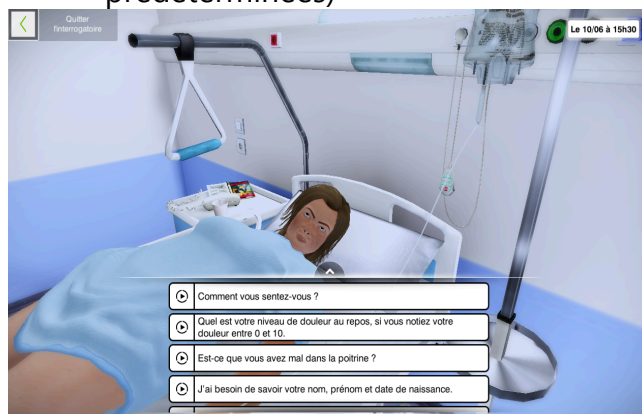
- c. Actions du joueur concernant l'examen clinique. Par exemple : prise de pression artérielle et surveillance du drain postopératoire.



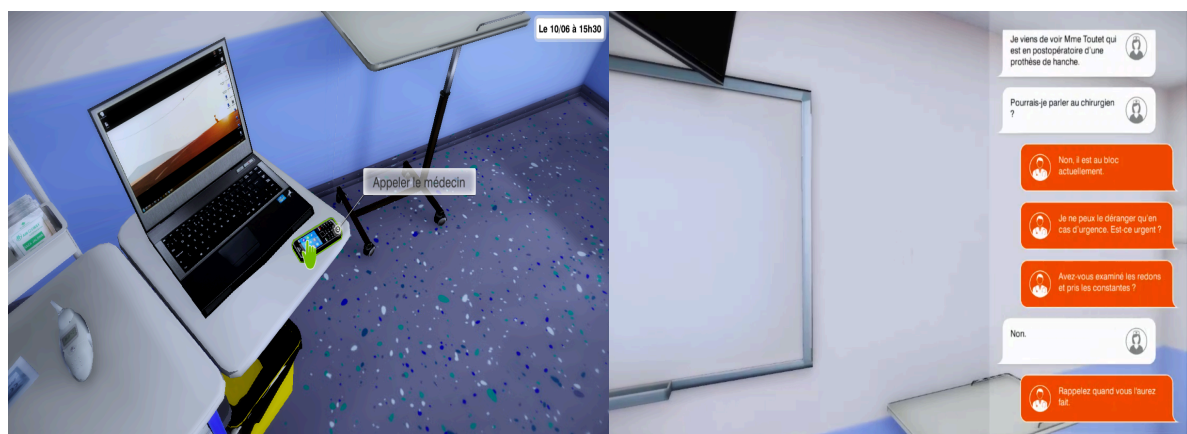
D. Dossier patient (prescription, transmission ...)

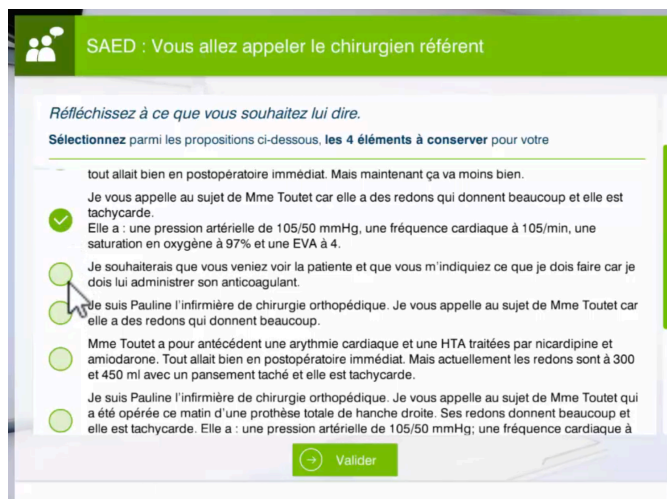


E. Interrogatoire du patient (par un menu déroulant avec des questions prédéterminées)

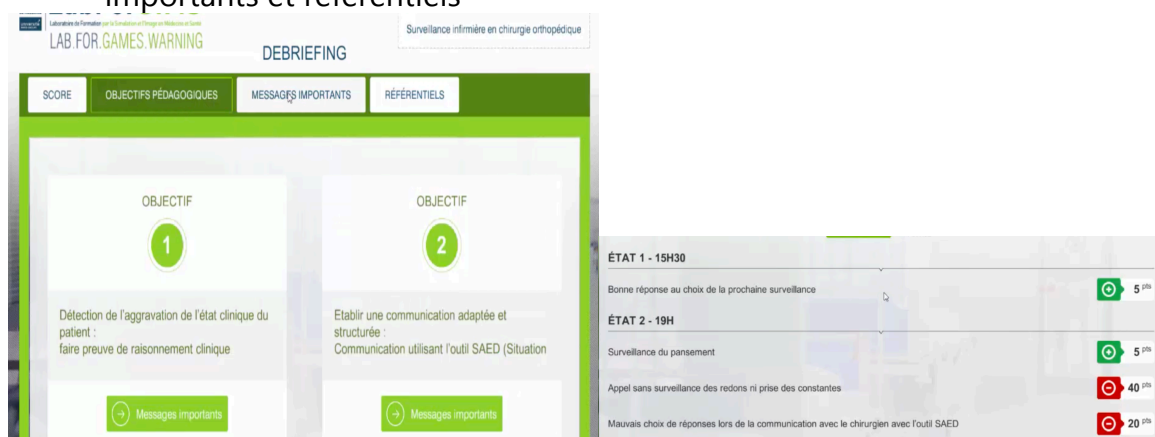


F. Appel du médecin (avec un téléphone et en utilisant la méthode SAED pour présenter le cas)





G. Feedback automatique avec le score, objectifs pédagogiques, messages importants et référentiels



Une fois *LabForGames Warning* développé, l'objectif de cette thèse a été l'évaluation expérimentale du raisonnement clinique dans le cadre de ce jeu sérieux pour la formation des professionnels de santé.

Une 1ère étude a évalué la validité du jeu sérieux *LabForGames Warning* selon la validité d'apparence, de contenu et du construit puis a été adaptée selon le cadre théorique proposé par Messick⁴⁶. Une 2nde étude a évalué l'efficacité de la simulation numérique avec *LabForGames Warning* sur l'apprentissage du raisonnement clinique en comparant un groupe d'étudiants infirmiers formé par le jeu sérieux par rapport à un groupe formé par un enseignement traditionnel.

3. Étude 1: Évaluation de la validité du jeu sérieux *LabForGames Warning*.

3.1. Introduction : les cadres théoriques de validité d'un outil pédagogique

Un outil pédagogique, comme le jeu sérieux, peut être utilisé à visée formative ou évaluative (sommative). A visée formative, il sera intégré lors de la formation pour permettre d'apporter ou améliorer un apprentissage. A visée évaluative, il sera utilisé en fin de formation pour évaluer l'apprenant. Mais avant d'utiliser un outil pédagogique, il est important d'en évaluer sa validité¹⁰⁶. La validation est un processus de recueil des preuves de validité pour interpréter de manière appropriée les résultats d'une évaluation⁴⁶. Plusieurs cadres théoriques de validité ont été décrits dans la littérature et plus particulièrement pour les jeux sérieux^{46,58,107}.

3.1.1. Cadre théorique classique de validité

Ce concept classique de validité est ancien mais reste encore couramment utilisé dans la littérature, incluant celle des jeux sérieux. Dans ce cadre théorique, la validité d'un outil de mesure inclut la validité du contenu, d'apparence, du construit, de concurrence et prédictive^{46,58,106,107}. La validité du contenu reflète la corrélation entre le contenu médical du jeu et l'objectif éducatif visé avec une évaluation positive et uniforme par un jury d'experts. La validité d'apparence reflète le niveau de ressemblance entre le contenu médical du jeu et la réalité, selon l'avis de débutants et d'experts. La validité du construit reflète la capacité de l'outil à mesurer une différence de performance entre des groupes n'ayant pas le même niveau de compétence dans la réalité (différence entre novice et expert). Le tableau 5 résume les définitions de chaque type de validité et des exemples ou analyses possibles (exemples de preuves) (Tableau 6).

Tableau 6. Définitions de chaque type de validité et exemples de preuves du cadre théorique classique de validité .

	Définitions	Exemples/analyses possibles
Validité de contenu <i>Content validity</i>	Reflète la corrélation entre le contenu médical du jeu et l'objectif éducatif visé avec une évaluation positive et uniforme par un jury d'experts en termes de contenu pédagogique	Évaluation positive et uniforme du contenu et des paramètres des tests par un jury d'experts pour sa légitimité. Par exemple : préparation par et accord entre les formateurs avec plusieurs tours de discussion
Validité d'apparence <i>= face validity = apparent validity</i>	Reflète le niveau de ressemblance entre le contenu médical du jeu et la réalité, selon l'avis de débutants et d'experts	Évaluation positive et uniforme par des utilisateurs novices et experts, sur la pertinence et ressemblance du jeu par rapport à la réalité

Validité du construit <i>Construct validity</i>	Reflète la capacité d'un instrument à mesurer une différence de performance entre des groupes n'ayant pas le même niveau de compétence dans la réalité.	Bon reflet du score sur les différents niveaux de compétence entre novices et experts Analyse factorielle Stabilité dans le temps
Validité de concurrence <i>Concurrent validity</i>	Concordance entre les résultats d'un test expérimental (le jeu sérieux) et un test de référence, censés évaluer la même démarche médicale	Bonne corrélation entre les résultats obtenus au test expérimental et au test de référence
Validité prédictive	Concordance entre le niveau de performance suggéré par le test et l'aptitude réelle, selon un système de notation	Bonne corrélation entre les résultats obtenus lors de l'entraînement au jeu et les aptitudes dans la vie réelle

Ce cadre théorique classique a été beaucoup utilisé pour décrire la validité d'un jeu sérieux. Dès 2012, l'équipe de Graafland et al. met en avant son importance et recommande l'utilisation de ce cadre théorique ^{58,107}. Depuis et encore à l'heure actuelle, de nombreuses équipes ont conduit des études de validation en utilisant ce cadre théorique ^{56,82,91,108-111}. Cependant, depuis une dizaine d'années, les sciences de l'éducation recommandent de remplacer ce cadre classique par le cadre théorique proposé par Messick et/ou Kane ^{46,106,112,113}.

3.1.2. Cadre théorique de preuve de validité proposé par Messick

Le cadre théorique de la preuve de validité a été proposé dès 1989 par Messick et adopté comme standard en 1999 et 2014 par *l'American Educational Research Association, American Psychological Association, National Council on Measurement in Education* ^{46,106,113}. Il comprend 5 preuves de validité : la preuve de contenu, la structure interne, la relation avec les autres variables, le processus de réponse, et la conséquence du test. Les définitions et exemples de preuves sont résumés dans le Tableau 7. On peut remarquer que certaines preuves correspondent à celles du cadre théorique classique décrit précédemment.

Tableau 7. Définitions et exemples du cadre théorique de preuve de validité proposé par Messick

Items de la preuve de validité	Définition	Exemples/analyses possibles
Preuve de contenu	Etre certain que le contenu du test reflète ce qu'il doit mesurer	Uniquement opinion des experts (Questionnaire avec interview, méthode Delphi)
Structure interne	Fiabilité du test	Fiabilité Inter-items (Cronbach alpha : coefficient de cohérence interne ; "acceptable" si $>0,7$) et/ou Fiabilité test-retest (Coefficient de corrélation entre 2 mesures séparées dans le temps et/ou inter-juge (Coefficient de corrélation entre 2 évaluateurs différents pour le même test)
Relation avec les autres variables	Comparer tests entre groupes de différentes compétences	Comparaison experts - novices
Processus de réponse	Contrôler la qualité de l'évaluation	Minimiser les biais de l'évaluation en analysant les évaluateurs et les examinés : Instruction écrite standardisée Évaluateurs en aveugle Analyser les examinés (vidéo, méthode pensée à haute voix)
Conséquence du test	Impact attendu du test	Établir un seuil du test (courbe ROC)

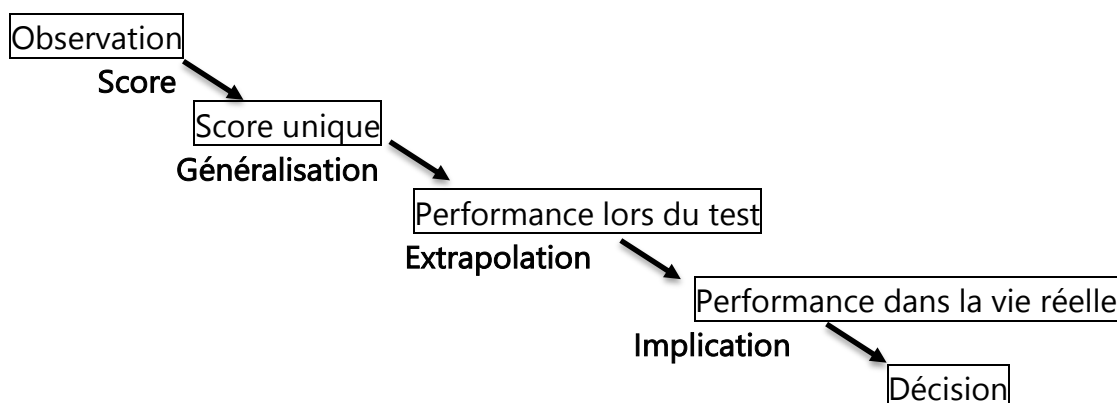
Bien que l'utilisation de ce « nouveau » cadre théorique soit recommandée^{46,106,112}, à ce jour, peu d'études dans le domaine de la simulation et notamment le jeu sérieux l'utilisent¹¹⁴. Dans leur revue systématique, Borgensen et al. ont rapporté que seulement 6,6% (33/498) des études en simulation chirurgicale publiées jusqu'en 2017 utilisaient le cadre de validité selon Messick¹¹⁴. De plus, 5 études seulement avaient évalué les cinq domaines dans les études chirurgicales incluses. Les autres études (93,4%) utilisaient l'ancien cadre théorique classique. En 2018, l'équipe de Gerard a utilisé ce cadre théorique pour établir les preuves de validation du jeu sérieux *PediatricSim* sur des urgences pédiatriques¹¹⁵.

3.1.3. Cadre théorique des inférences/déductions de validation proposé par Kane

Un autre cadre théorique de validité récent a été décrit par Kane^{116,117}. Il identifie 4 inférences clés dans le processus d'évaluation : le score, la généralisation, l'extrapolation et l'implication. L'ensemble de ce processus est décrit dans le Tableau 8. A notre connaissance, ce concept plus récent n'a pas encore été utilisé pour étudier la validité d'un jeu sérieux.

Tableau 8 Définitions et exemples de preuves du concept des inférences/déductions de validation de Kane

4 inférences clés de validation	Définition	Exemples/analyses possibles
Score	Sur la base d'une évaluation/observation, on obtient un score global ou une description narrative (synthèse) qui reflète la performance dans le cadre du test	Procédure de construction des items, options des réponses
Généralisation	L'ensemble des scores ou descriptions narratives (synthèse) reflète la performance dans le cadre d'un test	Taille de l'échantillon Fiabilité inter-items, et/ou fiabilité test-retest et/ou inter-juge
Extrapolation	Cette performance lors du test est supposée refléter la performance souhaitée en condition réelle	Comparaison novice et expert Analyse factorielle
Implication	La performance constitue une base rationnelle pour prendre une décision	Courbe ROC Conséquence score/examens



Au début de la construction de notre étude, le cadre théorique classique était le plus couramment utilisé dans la littérature des jeux sérieux^{56,82,110,111}. L'abandon de ce concept ancien étant fortement recommandé dans le domaine des sciences de l'éducation⁴⁶, nous avons choisi de présenter les résultats selon le cadre théorique proposée par Messick.

L'objectif de cette étude était d'évaluer les preuves de validité du jeu sérieux *LabForGames Warning* dédié à la détection de la dégradation d'un patient et la communication interprofessionnelle.

3.2. Article 1- Assessing validity evidence for a serious game dedicated to patient clinical deterioration and communication

Article accepté le 16/04/20 pour *Advances in Simulation*.

ASIM-D-19-00031R1

Assessing validity evidence for a serious game dedicated to patient clinical deterioration and communication.

Blanié Antonia M.D. ^{*(1,2,3)}, Amorim Michel-Ange Ph.D. ⁽³⁾, Meffert Arnaud ^(1,2), Perrot Corinne ⁽⁴⁾, Dondelli Lydie ⁽⁴⁾, Benhamou Dan M.D. ^(1,2,3)

** Corresponding Author: Dr Blanié, antonia.blanie@aphp.fr*

(1) Centre de simulation LabForSIMS, Faculté de médecine Paris Saclay, 94275 Le Kremlin Bicêtre, France. (2) Département d'Anesthésie-Réanimation chirurgicale, CHU Bicêtre, 94275 Le Kremlin Bicêtre, France. (3) CIAMS, Université Paris-Saclay, 91405 Orsay Cedex, France, and Université d'Orléans, 45067, Orléans, France. (4) IFSI Sud Francilien, 91106 Corbeil-Essonnes.

Abbreviated title: Assessing validity evidence for a serious game.

Abstract

Background: A serious game (SG) is a useful tool for nurse training. The objectives of this study were to assess validity evidence of a new SG designed to improve nurses' ability to detect patient clinical deterioration.

Methods: The SG (*LabForGames Warning*) was developed through interaction between clinical and pedagogical experts and one developer. For the game study, consenting nurses were divided into three groups: nursing students (pre-graduate) (Group S), recently graduated nurses (graduated <2 years before the study) (Group R) and expert nurses (graduated >4 years before the study and working in an ICU) (Group E). Each volunteer played three cases of the game (haemorrhage, brain trauma and obstructed intestinal tract). The validity evidence was assessed following Messick's framework: content, response process (questionnaire, observational analysis), internal structure, relations to other variables (by scoring each case and measuring playing time) and consequences (a posteriori analysis).

Results: The content validity was supported by the game design produced by clinical, pedagogical and interprofessional experts in accordance with the French nurse training curriculum, literature review and pilot testing. Seventy-one nurses participated in the study: S (n=25), R (n=25), E (n=21). The content validity in all three cases was highly valued by Group E. The response process evidence was supported by good security

control. There was no significant difference in the three groups' high rating of the game's realism, satisfaction and educational value. All participants stated that their knowledge of the different steps of the clinical reasoning process had improved. Regarding the internal structure, the factor analysis showed a common source of variance between the steps of the clinical reasoning process and communication or the situational awareness errors made predominantly by students. No statistical difference was observed between groups regarding scores and playing time. A posteriori analysis of the results of final examinations assessing study-related topics found no significant difference between Group S participants and students who did not participate in the study.

Conclusion: While it appears that this SG cannot be used for summative assessment (score validity undemonstrated), it is positively valued as an educational tool.

Trial registration: ClinicalTrials.gov ID: NCT03092440

Key words: serious game, simulation, validity evidence, patient deterioration

Background

Detection of patient deterioration is a major healthcare problem since a modification of physiological parameters often precedes acute patient clinical deterioration by 6 to 24 hours^{1 2 3}. The association of i) early detection ii) speed of response, and iii) quality of clinical response influences the patient prognosis. Many studies have shown that delayed diagnosis of an ongoing complication increases morbidity and mortality³. The education of nurses, who are frontline healthcare providers, is therefore essential.

When nurses are confronted with a case of clinical deterioration, they must not only recognize the incident but also notify the medical team. The use of a safe and standardized communication method such as the SBAR method^{4 5} improves patient safety^{6 7}. Training in understanding the role of appropriate communication and in the use of such a tool is therefore essential for healthcare professionals.

When compared with high-fidelity simulation, serious games (SG) possess an interesting immersive capacity and offer the advantage of training a large number of healthcare professionals in a limited amount of time using reduced educational resources^{8,9}. Moreover, SG are standardized cases providing automated feedback. SG can be used to develop both technical and non-technical skills¹⁰⁻¹². We developed a SG called *LabforGames Warning*, which aims to improve nursing students' interprofessional communication behaviour and their ability to detect patient clinical deterioration. Training in these essential skills will be soon added to the French nursing curriculum. In another study by our team, clinical reasoning was assessed in nursing students after a training course dedicated to the detection of patient deterioration, comparing a serious game-based simulation course with a traditional teaching course¹³. Although no significant educational difference was found between the two methods, participants reported greater satisfaction and motivation with serious game-based simulation training. However, the validity of this SG needed to be assessed before it could be used widely in professional healthcare education^{14,15}. The objective of this

study was to assess the validity evidence of *LabForGames Warning* before the game is used in educational activities.

Methods

SG development

The SG project was promoted by the Paris Sud university simulation centre ~~the~~ (LabForSIMS) in collaboration with four nursing schools (Sud Francilien, Perray Vaucluse, Paul Guiraud and Etampes) through a grant from the Ile-de-France Healthcare Regional Agency (ARS).

Three virtual clinical cases described below were developed through iterative dialogues between the pedagogical team and the developer (Interaction Healthcare®, Levallois-Perret, France). The medical instructors were clinical experts (teachers at four nursing schools and anaesthesiologists) and were also involved in the simulation centre. The educational objectives chosen for the SG were the detection of clinical deterioration and interprofessional communication. In the game, nurses are required to identify clinical deterioration in three different clinical situations and to notify the medical team accordingly based on the patient's clinical severity. *LabForGames Warning* derived its name from the early warning scoring system described in literature ¹⁶. As the SG focuses on nursing students, the objectives needed to conform to the French nurse training curriculum¹⁷.

In each clinical scenario, three consecutive steps (mildly abnormal, moderate aggravation and serious condition) were constructed to reproduce a specific complication of increasing severity in order to introduce the concept of early warning signs ¹⁶. The three cases were of equal moderate complexity. The clinical cases created were as follows:

Case 1 (post-operative haemorrhage): an adult female patient having undergone a scheduled total hip prosthesis earlier in the day and who is lying in her ward room bed immediately after arrival from the post-anaesthesia care unit. Post-operative haemorrhage from the surgical site is occurring progressively.

Case 2 (brain trauma): an elderly patient with dementia living in a nursing home whose anticoagulation is associated with progressively developing neurological deterioration following brain trauma from a fall.

Case 3 (obstructed intestinal tract): a schizophrenic patient hospitalized in a psychiatric ward with intestinal obstruction of progressively increasing severity.

Learning safe and standardized communication was an additional educational objective of the game ^{6,7}. We chose to train nursing students in the SBAR method, (Situation, Background, Assessment, Recommendation), which has been translated into French by the French Health Authority ⁵.

During the case, participants can perform different actions: history taking, clinical exams (circulatory assessment, neurologic assessment, skin temperature, etc.), care report writing and calling the physician. Screen shots of *LabForGames Warning* are provided in Figure 1 and Appendix 1.

At the end of each scenario, virtual automatic feedback was presented to the participant. Feedback included main guidelines and key messages about the detection of patient clinical deterioration (in general and in the specific case) and the SBAR method, as well as individualized global and detailed scoring (see Appendix 1, panel g, for an example).

The criteria for the detailed scoring had previously been established by the pedagogical team. The participant's clinical examination actions (checking arterial pressure, pain, etc.) and his/her decision (to call the physician, etc.) were assigned positive, negative or neutral points depending on the steps of the case. Moreover, positive or negative points were assigned to the quality of communication during the SBAR tool part of the game. The detailed score of Case 1 is presented in Appendix 2.

Study description

In this prospective, observational and non-interventional study, the participants were divided into three groups after giving informed consent.

Student nurse (S) group: graduate nursing students at the end of their second year of training.

Recently graduated (R) group: nurses having graduated less than two years before the study, who worked in a medical or surgical ward.

Expert nurse (E) group: nurses having graduated more than four years before the study, who worked in an intensive care unit.

The gaming sessions were held at the LabForSIMS simulation centre at the Paris Sud medical school and at the Sud Francilien nursing school. Each volunteer played Cases 1, 2 and 3 in randomized order on an individual computer.

Validity evidence

The objective of this study was to assess the validity evidence of *LabForGames Warning* before using the game in educational activities.

At the beginning of our study and according to Graafland et al., the validity of a SG should be assessed by using content validity, face validity, construct validity, concurrent validity and predictive validity^{15 10 18}. However, this classical validation framework may be replaced by those of Messick or Kane^{14,19}. To date, few studies in the simulation field have used the latter frameworks¹⁹⁻²². In their systematic review, Borgensen et al. reported that only 6.6% of the surgical simulation studies published up to 2017 used Messick's recommended validity framework²¹. Moreover, only five studies have assessed all five domains of the Messick framework in the surgical studies reviewed. In the present study, the following five domains of Messick's framework for validity evidence were assessed: content, response process, internal structure, relations to other variables and consequences¹⁴.

Content is defined by "the relationship between the content of a test and the construct it is intended to measure"¹⁴. The educational content, learning objectives and branched steps were developed by clinical and pedagogical experts (nine instructors of four nursing schools and three anaesthesiologists who were also simulation

instructors) in conformity with the French nurse training curriculum ⁵ and literature review. For each scenario, the script, pedagogical objectives, feedback and scoring were written, reviewed and validated through expert consensus. Virtual clinical case development was also the product of iterative dialogues between the pedagogical team and the developer. Pilot testing involved pedagogical clinical experts (different from Group E) and corrections were made before the final version was used in the study. Moreover, content validity in the study was assessed by expert nurses (Group E) who judged the medical content and the educational objectives of the game (using a ten-point Likert scale).

The response process is "the fit between the construct and the detailed nature of performance [...] actually engaged in" ¹⁴. During the SG sessions, we controlled the security (defined as the prevention of cheating) ¹⁴ and the quality of this assessment. All participants completed a standardized tutorial just prior to using the SG. Each participant played the game on an individual computer with no personal documents. An instructor was present at all times to prevent cheating. The instructors had no access to the scores. We also analysed the participants' perception with the aid of a questionnaire at the end of the SG session. The following participant characteristics—sex, age, post-graduate experience, intensive care experience and previous video gaming activity (entertainment and professional education)—were recorded (Table 1). This questionnaire also assessed the participants' perception of three main themes: satisfaction with the educational tool, game realism and future professional impact (using a ten-point Likert scale) (Table 2). Self-assessment of the clinical reasoning learning process was also recorded after the session. This questionnaire, translated into French, had previously been related by Koivisto et al. and assesses the various steps of clinical nursing reasoning as defined by Levett-Jones et al. ^{23 24}. Each question assesses a specific step in the clinical reasoning process ("I learned to...") with the use of a five-point Likert scale. The global result (graded out of 70) was obtained by totalling the values assigned to the 14 questions (Table 3).

The internal structure is defined by "the relationship among data items within the assessment and how these relate to the overarching construct" ¹⁴. A factor analysis (Principal Component Analysis) was used to identify the relations between the main steps of the clinical reasoning process (using the data from the self-assessment questionnaire presented in Table 3) ^{23,24} and the non-technical errors (situational awareness, communication) at each level of expertise (S, R, E groups). Concerning errors, negative points were classified as situational awareness errors when they related to the diagnostic part of the scenario, and as communication errors when they occurred during the SBAR tool part of the game.

Relations to other variables are the "degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" ¹⁴. The ability of this SG to measure differences between groups of different skill levels was assessed by comparing the scores and the playing time of Groups S, R and E. The scores obtained for each case were graded out of 100 points. The playing time in each case (in minutes) was also assessed.

Consequences are “the impact, beneficial or harmful and intended or unintended, of assessment”¹⁴. A posteriori, we identified the results of examinations related to training sequences that were associated with the SG pedagogical objectives: “Care project module”; “Emergency module”; and “Plan and implement nursing interventions and therapeutics module”. We then compared the exam results obtained by Group S participants and those of students who had not participated in the study (i.e. the remaining students in the same class who did not participate in the SG session).

Statistical analysis

Game scores were used to define the number of participants to be included. Considering that Group S would obtain a novice score (no reference available but estimated at 60/100) and that Group E would have an approximate score of 80/100 (no reference available), the difference between the students and the experts was 20/100. Considering a standard deviation of 15 points, the sample size was 12 per group with the use of a two-tailed analysis (alpha risk= 0.05 and power of 0.9)²⁵. In view of the risk of attrition, we decided to form groups of 20 participants.

The results are presented as means $\pm$ standard deviation or percentages and confidence intervals. After the normal distribution assessment, statistical analysis was performed using parametric tests (one-way ANOVA test or Chi2 test, followed by post hoc tests in the case of significant comparison) (JMP software, SAS Institute ®). The factorial analysis (Principal Component Analysis) was performed using Statistica software (StatSoft Inc. ®). A *p* value less than 0.05 was considered significant and adjustment for multiple comparisons was performed.

Ethical statement

This study was approved (on March 30, 2017) by The Institutional Review Board of Paris Saclay University (CERNI). The project has been registered on ClinicalTrials.gov (ClinicalTrials.gov ID: NCT03092440)²⁶. The study was conducted with the use of the CONSORT tool adapted for simulation studies and the GREET Tool for educational studies²⁷.

Results

Inclusion

Seventy-one nurses and nursing students participated in this study voluntarily between March and September 2017. Participants in Group S were students at the Sud Francilien nursing school, whereas graduated nurses were recruited at the Kremlin Bicêtre University teaching hospital (Group R from medical and surgical units and Group E from two ICUs). Participant characteristics are presented in Table 1. One student experienced a technical problem during Case 1 so no data could be stored for the analysis of Case 1. Another student failed to record the clinical reasoning self-assessment. All of the participants played all three cases to the end.

Content evidence

The nurses in Group E considered the SG as providing complete and good nursing care regarding the medical content and educational objectives for the three cases (Q1) (Table 2). The global educational value of this SG was also positively perceived by Group E (Q8-9).

Response process evidence

A summary of the perception survey is shown in Table 2. All three groups scored the realism and graphics of the three scenarios positively with no significant difference between the groups (Q2-Q5, Q7). Group E considered the care record (Q6) less realistic than did the other two groups ($p < 0.05$).

The global educational value of the SG was perceived positively with no significant difference by all three groups (Q8-9). Groups S and R declared that the game could improve their skills (Q10) and could have an impact on their professional work (Q11). Conversely, Group E perceived the game as less useful in improving their practice ($p < 0.05$) (Q10-11). However, all three groups stated they would recommend this session to students or colleagues (Q12).

Following training with the SG, all participants considered that their knowledge of the different steps of the clinical reasoning process had increased (self-assessment). There was no significant difference in the Group Scores (Table 3).

Internal structure evidence: Factor analysis

Factor analysis was used to confirm the validity of the self-reporting questionnaire and to distinguish between the factors studied (realism, educational content and impact on the participant) (Appendix 3 b). Factor analysis was also used to identify relations between the clinical reasoning process and errors (communication and situational awareness) in the groups (Figure 2 and Appendix 3 c). In Group S, the first part of clinical reasoning (collect/process/identify) was linked to both situational awareness and communication errors whereas the implementation part (establish goal/take action) was linked to communication errors only. In Group R, only communication was found to be related to the first part of reasoning (identify) on the one hand, and the implementation part (decision/treatment) on the other. In Group E, no relation could be observed between clinical reasoning and errors of communication or situational awareness.

Evidence regarding relations to other variables: comparison of scores and playing time between groups

There was no significant difference in scores between groups (main outcome) and no significant difference was found in the playing time between groups (Table 4). Moreover, no correlation between individual scores and playing time was observed between groups or for the whole set of participants (Case 1: $r = -0.08$; $p=0.48$; Case 2: $r = 0.06$, $p=0.61$; Case 3: $r = -0.10$, $p=0.43$); nor did factor analysis demonstrate any relationship between the scores and participants' experience (Appendix 1). Moreover,

no relationship was observed between the scores and questions about content and face validity (Appendix 3a).

Evidence for consequences

A posteriori analysis demonstrated no significant difference in the three exam results (graded /20) between Group S (n=25) and the control group (n=111) (Care project module: 14.3 ± 2.4 vs 13.9 ± 2.3 , $p=0.41$; Emergency module: 10.8 ± 1.8 vs 10.6 ± 2.3 , $p=0.68$; and Plan and implement nursing interventions and therapeutics module: 13.8 ± 3.4 vs 12.4 ± 3.1 , $p=0.07$, respectively).

Discussion

Validity assessment is necessary for an SG, as for any new educational tool^{14,15}. In this study, we used the five domains of validity evidence described by Messick et al.¹⁴ (content, response process, internal structure, relations to other variables and consequences). The main findings are that neither the gameplay scores nor the playing time of *LabForGames Warning* differentiated the level of the nurses' skills. However, other domains of validity evidence for this SG were demonstrated.

First, content validity evidence is the most frequently assessed domain in the educational literature^{14 28 29 30 14,31}. *LabForGames Warning* (educational content and objectives, different branched steps, scoring) was produced by clinical, pedagogical and interprofessional experts in conformity with the French nurse training curriculum⁵ and literature review. Effective educational content was demonstrated as experts (Group E) expressed a positive attitude toward the medical algorithm and the nurse decision-making process, confirming content legitimacy.

A second domain of validity evidence was the response process, which was assessed using rigorous quality and security control during the study. Moreover, both experts and novices were asked to assess the tool's apparent similarity with reality and its usefulness for educational purposes. Evaluation by experts is especially crucial in order to collect validity evidence. In our study, the experts were from the units in which the game's cases took place (orthopaedic department and psychiatry department but not from nursing home). Moreover, nurses work in many different units (surgery, medicine, etc.) prior to graduating. Realism was considered for the whole gameplay but also for its different parts (i.e. nursing care, clinical examination, care records and graphics). The difference found for care record realism between groups may be explained by the fact that electronic care records are not available in all hospital units, which complicates extrapolation. Moreover, the SG's ability to improve skills, or the impact on the professional outcome, were evaluated positively, especially by students and recently graduated nurses, confirming our initial educational choice to target this population. Satisfaction with the training process, skill improvement self-assessment and the impact on professional outcomes were considered satisfactory. Moreover, after training with the SG, all of the participants felt that their skills had improved in the

different steps of the nurse clinical reasoning process, with a global score of 52/70. Teaching clinical reasoning with the aid of an SG appears to be of value and relevant for trainees. The virtual cases represent experiential learning as described by Kolb³² and explore the four domains of the clinical reasoning process³³. Learning of clinical reasoning is complex to assess³⁴ and although self-assessment involves only a subjective perception, it does provide important information. The tool we used was based on the clinical reasoning process described by Levett-Jones and used by Koivisto^{24 23}. Other tools have also been published^{35 36}. Despite their uncertain validity, these tools aim to assess the various steps of clinical reasoning. However, most studies have analysed only the results of the overall reasoning process (i.e. diagnosis and treatment) but not all of the steps of clinical reasoning^{37-40 8}.

Third, with regard to validity evidence for internal structure, factor analysis appeared to be a useful tool to identify behaviours specific to each group by assessing the relations between parts of the clinical reasoning process and errors. Clinical reasoning is a complex cognitive process⁴¹. According to the dual process theory, two cognitive systems are used by healthcare providers. System 1 is heuristic reasoning based on illness pattern recognition (matching an actual configuration of signs with previously encountered equivalent situations), allowing intuitive mental shortcuts to reduce the cognitive load of decision making. System 2 is an analytical reasoning model that integrates all available information and requires great effort. Simply stated, System 1 is more easily implemented by experts due to clinical experience whereas System 2 would be used more often by novices. Interestingly in our study, organisation of links between parts of the clinical reasoning process and errors was found to be an indicator of expertise. In Group E, no relation between clinical reasoning, communication and situation awareness was observed, suggesting that encapsulation of clinical reasoning occurs with experience and is congruent with a more frequent use of System 1 (intuitive) processing⁴¹. Each step is dependent on the following one, as a unique "module" of clinical reasoning⁴². Moreover, the independence of the clinical reasoning items of the self-assessment suggests that the modularity of clinical reasoning is embedded in a deeper structure that is inaccessible to awareness and with no explicit link. On the contrary, in Group S, the first part of reasoning was linked to both situational awareness and communication errors whereas the implementation part was linked to communication errors only. In Group R, only communication was found to be related positively to the first part of reasoning, and negatively to the implementation part, which suggests a beginning of expertise. However, although we did not assess situation awareness itself using validated methods⁴³, we classified errors (negative points) occurring in the diagnostic part of the scenario and those related to decisions regarding the next monitoring interval as "situation awareness" errors, and investigated relations with other variables through factor analysis. Moreover, we did not record the exact time at which each action was done during each step. Furthermore, we do not know precisely when deterioration was identified since situational awareness is a progressive process with interconnecting steps. Only the global playtime of cases was

recorded with no significant difference between groups. Therefore, individualization of the different steps of the clinical reasoning process during the game was not possible and the manner in which participants performed individually in each case could not be determined. It could be interesting to introduce markers for each step of the clinical reasoning process in a future version of the game.

The fourth domain of validity evidence studied was the relation to other variables by comparing the groups' scores and playing time. In previous studies, the validity of the scoring system was either not assessed^{24,44-46} or was assessed only by means of "static" multiple choice questions that were obvious to the participant since they represented the logical steps of a surgical procedure^{29 30 28}. In contrast, our game design was branched and included algorithms with many possibilities and different interactions between the patient and the physician, as the various ramifications in the storyboard aimed to reproduce the most likely clinical situations. Our SG combined the assessment of several non-technical skills, including situational awareness and communication. Therefore, our results highlight the difficulty in establishing a scoring system due to several interactive and complex problems. In a similar SG in critical paediatric emergency care, Gerard et al. demonstrated validity evidence for *Pediatric Sim* Game scores with higher scores for attendings followed by residents, than for medical students²⁰. A strong positive correlation was found between game scores and written knowledge test scores. However, as with our SG (Table 4), game scores were low across all groups (68/100 for attendings), which confirms the difficulty in constructing a score. Additionally, the score explores only a limited part of the tool and not all of its pedagogical impact and utility. The assessment of construct validity is essential if the SG is to be used in a summative educational process. Indeed, if the score's construct validity is not demonstrated, the game cannot be used to evaluate student learning at the end of an instructional unit. To our knowledge, no SG like this one has been used for summative assessment. However, this version of our SG can be used for training in an educational program since some domains of validity evidence could be demonstrated. Certain teams have already included an SG in their training program^{44 24 47}. The nursing school participating in this project recently introduced it in the student curriculum because detailed scoring analysis can improve the instructors' debriefing since they can use the detailed scoring of each trainee, available on an e-platform. More studies are necessary to define the place of this tool in professional healthcare education.

Since *LabForGames Warning* scores could not differentiate between the levels of expertise, one might wonder how it might be improved. The instructors tried to align the scoring system to case complexity, for which a certain level of proficiency was expected. In the post-operative haemorrhage case, for example, analysis of the detailed subscores showed that the majority (>85%) of essential nursing care actions were performed by each group. Participant actions were stereotyped and limited. Allocating points in a different manner and/or increasing the number of tasks available to the participant including some unnecessary (or even deceptive) actions might be more

discriminant. Indeed, one limitation of the game itself is that if the participant performs all actions, many positive points may be earned with no actual clinical reasoning. Recording the response time at each point could also be useful.

Playing time is a surrogate marker of the time it takes to care for the patient, collect data, make a decision and call for help. Although one might expect playing time to be longer for the novice than for the expert, this study found no difference in the playing time between the groups. Results of previous studies are mixed on this subject and do not consistently show a direct correlation between playing time and expertise^{28,48}. The absence of differences in playing time could be explained by the fact that the time devoted to each participant's action and communication are limited and predefined by the game itself.

When trying to study the consequences of using the SG (i.e. the last domain of validity evidence), a posteriori analysis found no significant difference in examination results between the student group having played the SG and the control group. However, no definitive conclusion can be drawn since the groups studied were not randomized.

One limitation of the study was that several items of our set of measures to assess validity were based on participant perception, although objective measures would appear more potent. Perception, however, is more often studied in the literature. For example, Graafland et al. and Sugand et al. validated the content and face validity of their SG with a self-report questionnaire^{28 29 30 31}.

Another limitation was that mean scores were low (< 50/100). Some negative points attached to the communication part were based on the use of the SBAR tool as even our expert nurses had received no previous training on how to use this tool for which a French version had only recently been made available⁵. However, even when scores were recalculated after excluding the SBAR tool, no significant differences were observed between groups.

Writing gameplay is a difficult task since each clinical situation has several possible outcome branches and poor writing may lead to low score results and poor discrimination. However, the pedagogical team was composed of medical experts (anaesthesiologists) and nursing experts, who were also experts in pedagogy. Interestingly, all of the teachers were also experienced high-fidelity session instructors.

Conclusions

In conclusion, our study demonstrated that the scores and the playing time of the game *LabForGames Warning* did not differentiate nurses' levels of clinical skills. However, validity evidence was obtained from the content, the response process and the internal structure. Although the present version cannot be used for the summative assessment of nursing students, our study has shown that this SG was well received by the participants and that it can be used for training in an educational program. More studies are necessary to improve SG scoring details, and, more generally speaking, to define this new tool's place in the field of education.

Declarations

Abbreviations

S: Student nurse

R: Recently graduated

E: Expert nurse

SG: Serious Game

Ethics approval and consent to participate

This study was approved (March 30, 2017) by The Institutional Review Board of Paris Saclay University (CERNI).

Consent for publication

The content of this paper is not being submitted elsewhere. All authors acknowledge their familiarity with these instructions and agree to the contents of the submitted paper. The manuscript has been read and approved by all co-authors.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare no conflict of interest.

Funding

Agence Régionale de Santé Ile de France (French Regional Authority for Health Care Regulation) which provided a grant to create the Serious Game and principally funded the development of the game in 2014 (before this research). The "Agence Régionale de Santé Ile de France" did not provide a grant for the research project. None of the authors received any personal financial support.

Authors' contributions

AB performed data collection and analysis, and interpreted the data regarding the Serious Game and was a major contributor to the writing of the manuscript.

MAA analysed and interpreted the data regarding the Serious Game and contributed to the writing of the manuscript.

AM performed data collection and analysis, and interpreted the data regarding the Serious Game.

CP and LD performed data collection.

DB analysed and interpreted the data regarding the Serious Game and contributed to the writing of the manuscript.

All authors have read and approved the final manuscript.

Acknowledgments

We would like to acknowledge the *LabForGames Warning* project team: Philippe Roulleau, Bonga Barcello De Carvalho, Hélène Bertrand, Alain Cazin, Véronique Mahon, Lionel Henriques, Nathalie Léon, Vincent Lebreton, Laure Legoff, Aurélie Woda, Bertrand Bech and Alexandre Renault.

References

1. Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation*. 2004;62(2):137-141.
2. Hillman KM, Bristow PJ, Chey T, et al. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med*. 2002;28(11):1629-1634.
3. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg*. 2009;250(6):1029-1034.
4. De Meester K, Verspuy M, Monsieurs KG, Van Bogaert P. SBAR improves nurse-physician communication and reduces unexpected death: a pre and post intervention study. *Resuscitation*. 2013;84(9):1192-1196.
5. http://www.has-sante.fr/portail/jcms/c_1776178/fr/saed-un-guide-pour-faciliter-la-communication-entre-professionnels-de-sante.
6. Nagpal K, Arora S, Vats A, et al. Failures in communication and information transfer across the surgical care pathway: interview study. *BMJ Qual Saf*. 2012;21(10):843-849.
7. Mackintosh N, Sandall J. Overcoming gendered and professional hierarchies in order to facilitate escalation of care in emergency situations: the role of standardised communication protocols. *Soc Sci Med*. 2010;71(9):1683-1686.
8. Liaw SY, Chan SW, Chen FG, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. *J Med Internet Res*. 2014;16(9):e214.
9. Haerling KA. Cost-Utility Analysis of Virtual and Mannequin-Based Simulation. *Simul Healthc*. 2018;13(1):33-40.
10. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg*. 2012;99(10):1322-1330.
11. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. *J Med Internet Res*. 2019;21(3):e12994.
12. Gorbanev I, Agudelo-Londono S, Gonzalez RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online*. 2018;23(1):1438718.
13. Blanie A, Amorim MA, Benhamou D. 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 Med Educ*. 2020;20(1):53.
14. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1:31.
15. Graafland M, Dankbaar M, Mert A, et al. How to systematically assess serious games applied to health care. *JMIR Serious Games*. 2014;2(2):e11.
16. Gao H, McDonnell A, Harrison DA, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*. 2007;33(4):667-679.

17. Legifrance.gouv.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037218115&categorieLien=id>. Accessed 03-17, 2020.
18. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872-883.
19. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978-988.
20. Gerard JM, Scalzo AJ, Borgman MA, et al. Validity Evidence for a Serious Game to Assess Performance on Critical Pediatric Emergency Medicine Scenarios. *Simul Healthc*. 2018;13(3):168-180.
21. Borgersen NJ, Naur TMH, Sorensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Ann Surg*. 2018;267(6):1063-1068.
22. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*. 2014;19(2):233-250.
23. Levett-Jones T, Hoffman K, Dempsey J, et al. The 'five rights' of clinical reasoning: an educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. *Nurse Educ Today*. 2010;30(6):515-520.
24. Koivisto JM, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: A cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. *Nurse Educ Today*. 2016;45:22-28.
25. Biostatgv. <https://marne.u707.jussieu.fr/biostatgv/>. Accessed 03-17, 2020.
26. Cheng A, Kessler D, Mackinnon R, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Adv Simul (Lond)*. 2016;1:25.
27. Phillips AC, Lewis LK, McEvoy MP, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). *BMC Med Educ*. 2016;16(1):237.
28. Graafland M, Vollebergh MF, Lagarde SM, van Haperen M, Bemelman WA, Schijven MP. A serious game can be a valid method to train clinical decision-making in surgery. *World J Surg*. 2014;38(12):3056-3062.
29. Sugand K, Mawkin M, Gupte C. Training effect of using Touch Surgery for intramedullary femoral nailing. *Injury*. 2016;47(2):448-452.
30. Kowalewski KF, Hendrie JD, Schmidt MW, et al. Validation of the mobile serious game application Touch Surgery for cognitive training and assessment of laparoscopic cholecystectomy. *Surg Endosc*. 2017;31(10):4058-4066.
31. Graafland M, Bemelman WA, Schijven MP. Appraisal of face and content validity of a serious game improving situational awareness in surgical training. *J Laparoendosc Adv Surg Tech A*. 2015;25(1):43-49.
32. Rouse DN. Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspect Health Inf Manag*. 2011;8:1c.
33. Cutrer WB, Sullivan WM, Fleming AE. Educational strategies for improving clinical reasoning. *Curr Probl Pediatr Adolesc Health Care*. 2013;43(9):248-257.
34. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ*. 2012;46(5):454-463.
35. Liou SR, Liu HC, Tsai HM, et al. The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. *J Adv Nurs*. 2016;72(3):707-717.
36. Liaw SY, Rashasegaran A, Wong LF, et al. Development and psychometric testing of a Clinical Reasoning Evaluation Simulation Tool (CREST) for assessing nursing students' abilities to recognize and respond to clinical deterioration. *Nurse Educ Today*. 2018;62:74-79.
37. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med*. 2010;85(10):1589-1602.

38. Knight JF, Carley S, Tregunna B, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. *Resuscitation*. 2010;81(9):1175-1179.
39. Kleinert R, Heiermann N, Plum PS, et al. Web-Based Immersive Virtual Patient Simulators: Positive Effect on Clinical Reasoning in Medical Education. *J Med Internet Res*. 2015;17(11):e263.
40. Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebo A. Teaching clinical reasoning and decision-making skills to nursing students: Design, development, and usability evaluation of a serious game. *Int J Med Inform*. 2016;94:39-48.
41. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011;16.
42. Fodor JA. *The Modularity of Mind*. MIT Press/Bradford Books; 1983.
43. Wright MC, Taekman JM, Endsley MR. Objective measures of situation awareness in a simulated medical environment. *Qual Saf Health Care*. 2004;13 Suppl 1:i65-71.
44. Bogossian FE, Cooper SJ, Cant R, Porter J, Forbes H, Team FAR. A trial of e-simulation of sudden patient deterioration (FIRST2ACT WEB) on student learning. *Nurse Educ Today*. 2015;35(10):e36-42.
45. Liaw SY, Wong LF, Lim EY, et al. Effectiveness of a Web-Based Simulation in Improving Nurses' Workplace Practice With Deteriorating Ward Patients: A Pre- and Postintervention Study. *J Med Internet Res*. 2016;18(2):e37.
46. Liaw SY, Chng DYJ, Wong LF, et al. The impact of a Web-based educational program on the recognition and management of deteriorating patients. *J Clin Nurs*. 2017;26(23-24):4848-4856.
47. Mohan D, Farris C, Fischhoff B, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ*. 2017;359:j5416.
48. Chaudery M, Clark J, Dafydd DA, et al. The Face, Content, and Construct Validity Assessment of a Focused Assessment in Sonography for Trauma Simulator. *J Surg Educ*. 2015;72(5):1032-1038.

Figure 1: Screen shots of *LabForGames Warning*



Figure 2: Links between errors (communication and situational awareness) and the clinical reasoning process as demonstrated by principal component factor analysis.

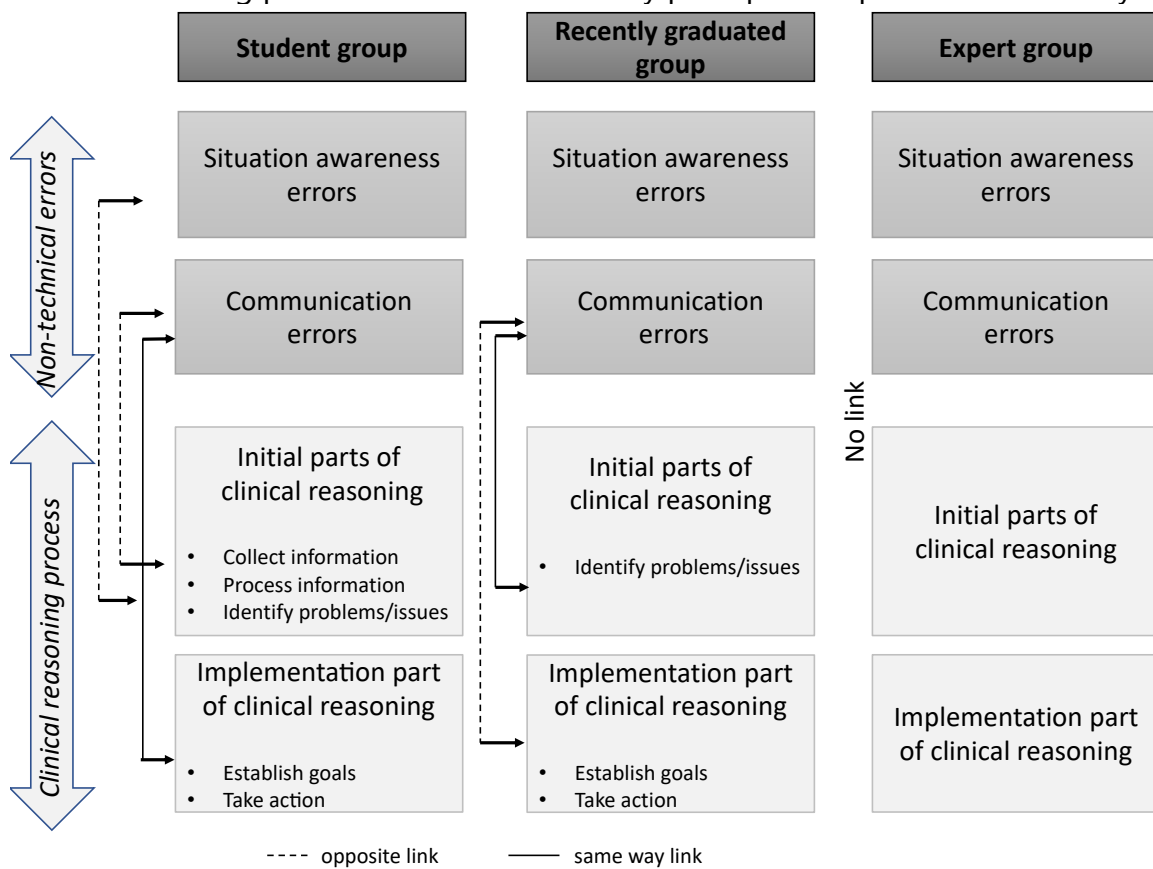


Table 1: Characteristics of players included in Group S (nursing students); Group R (recently graduated nurses) and Group E (expert nurses).

Results are reported as mean $\pm$ SD or as a percentage, and compared using ANOVA test or Chi2 test (followed by post hoc tests when significant).

*: p value < 0.05 between Groups S, R and E.

** : Group E $\neq$ S with $p < 0.05^{***}$: Group E $\neq$ Group R with $p < 0.05^{***}$: Group S $\neq$ Group R with $p < 0.05$.

\$: no response for one student.

	Group S n=25	Group R n=25	Group E n=21	p
Age (years)	25.0 $\pm$ 6.6	25.0 $\pm$ 4.2	31.6 $\pm$ 5.9 ^{**,***}	0.0004*
Sex F/M	22(88%)/3(12%)	19(76%)/6(24%)	14(67%)/7(33%)	0.22
Post-graduate professional experience (years)	0	1.4 $\pm$ 0.5	7.9 $\pm$ 4.5 ^{**,***}	0.0004*
Previous ICU experience (years)	0	0.04 $\pm$ 0.2	5.5 $\pm$ 3.9 ^{**,***}	
Video gaming activity \$: never 1 /month 1 / week Every day	17(71%) 5(21%) 2(8%) 0(0%)	16(64%) 5(20%) 1(4%) 3(12%)	14(66.7%) 5(24%) 2(10%) 0(0%)	0.40
Video gaming activity in healthcare (% of players)	1(4%)	0(0%)	4(19%) ^{**,***}	0.03*
Experience in specific situations (1 [none] to 10 [expert])				
- post-operative haemorrhage	2.1 $\pm$ 1.7	2.8 $\pm$ 2.3	7.3 $\pm$ 1.5 ^{**,***}	<.0001*
- brain trauma	2.1 $\pm$ 1.7	3.3 $\pm$ 2.6 ^{****}	7.8 $\pm$ 1.4 ^{**,***}	<.0001*
- intestinal obstruction	2.3 $\pm$ 1.8	2.6 $\pm$ 2.1	6.7 $\pm$ 1.9 ^{**,***}	<.0001*

Table 2: Results of the self-questionnaire assessing response process validity evidence in the three groups.

Results are described as means $\pm$ SD and compared using ANOVA (followed by post hoc comparison when significant). **: Group E $\neq$ S with $p < 0.05$; ***: Group E $\neq$ Group R with $p < 0.05$, ****: Group S $\neq$ Group R with $p < 0.05$.

	Group S n=25	Group R n=25	Group E n=21	p
Q1- Do you think that this serious game provided a complete and good nursing care experience with respect to the medical content and educational objectives? 1 (No, not at all) to 10 (completely correct and good)				
- post-operative haemorrhage scenario	7.8 $\pm$ 1.2	8.0 $\pm$ 1.3	7.8 $\pm$ 1.3	0.83
- brain trauma scenario	7.8 $\pm$ 1.3	7.0 $\pm$ 1.3	8.2 $\pm$ 1.0	0.48
- obstructed intestinal tract scenario	7.5 $\pm$ 1.4	7.7 $\pm$ 1.5	7.5 $\pm$ 1.4	0.86
Q2- Do you think that this serious game was realistic? 1 (not realistic) to 10 (very realistic)	7.7 $\pm$ 1.4	7.0 $\pm$ 1.6	7.5 $\pm$ 1.4	0.25
Q3- Do you think that the clinical scenario was realistic? 1 (not realistic) to 10 (very realistic)				
- post-operative haemorrhage scenario	8.0 $\pm$ 2.0	7.7 $\pm$ 1.6	8.2 $\pm$ 1.1	0.50
- brain trauma scenario	8.1 $\pm$ 1.4	7.8 $\pm$ 1.3	8.6 $\pm$ 0.7	0.07
- obstructed intestinal tract scenario	7.4 $\pm$ 2.0	7.6 $\pm$ 1.4	8.0 $\pm$ 1.1	0.39
Q4- Do you think that the history-taking was realistic? 1 (not realistic) to 10 (very realistic)	7.9 $\pm$ 1.3	7.4 $\pm$ 1.4	7.5 $\pm$ 1.2	0.41
Q5- Do you think that the clinical exam was realistic? 1 (not realistic) to 10 (very realistic)				
- post-operative haemorrhage scenario	7.8 $\pm$ 1.2	7.7 $\pm$ 1.4	8.1 $\pm$ 1.0	0.53
- brain trauma scenario	8.0 $\pm$ 1.0	7.8 $\pm$ 1.3	8.1 $\pm$ 1	0.58
- obstructed intestinal tract scenario	7.5 $\pm$ 1.5	7.8 $\pm$ 1.3	7.8 $\pm$ 1.1	0.76
Q6- Do you think that the care report was realistic? 1 (not realistic) to 10 (very realistic)	8.5 $\pm$ 1.0 ^{**,****}	7.5 $\pm$ 1.6	7.4 $\pm$ 1.9	0.02 [*]
Q7- Do you think that the graphics and animations were realistic? 1 (not realistic) to 10 (very realistic)	8.4 $\pm$ 0.9	7.6 $\pm$ 1.3	7.8 $\pm$ 1.7	0.06
Q8- Overall, are you satisfied with this serious game as an educational tool? 1 (not satisfied) to 10 (very satisfied)	7.0 $\pm$ 1.5	7.0 $\pm$ 1.7	7.0 $\pm$ 1.5	0.98
Q9- Do you think that incorporating the serious game concept improves your training? 1 (No, not at all) to 10 (Yes, very much)	8.4 $\pm$ 1.2	7.7 $\pm$ 2	6.9 $\pm$ 2.9	0.05

Q10- Do you think that this tool increases your skills? 1 (No, not at all) to 10 (Yes, very much)	8.4 ± 1.5	7.7 ± 1.8	5.5 ± 2.1 ^{**,***}	<.0001 [*]
Q11- Do you think that this training will have an impact on your professional work? 1 (No, not at all) to 10 (Yes, very much)	7.3 ± 1.3	6.8 ± 1.8	4.4 ± 1.9 ^{**,***}	<.0001 [*]
Q12- Would you recommend this training to other students or colleagues? 1 (No, not at all) to 10 (Yes, very much)	8.8 ± 1.2	8.2 ± 1.5	8.1 ± 1.8	0.17

Table 3: Results of self-assessment of learning the clinical reasoning process between groups.

Results are described by means $\pm$ SD and compared using ANOVA test (followed by post hoc comparison when significant). [£] Bonferroni criterion: $\alpha/14 = 0.0035$ to be significative.

One student was excluded for not answering the clinical reasoning self-questionnaire. The questionnaire, translated in French, describes the steps of clinical nursing reasoning as defined previously ^{20 43}

I learned: 1 (not at all) to 5 (very much)		Group S n=24	Group R n=25	Group E n=21	p
To collect information	Collect information by interviewing patient	4 $\pm$ 0.6	4.2 $\pm$ 0.7	4 $\pm$ 0.8	0.57
	Collect information by observing patient	3.7 $\pm$ 0.7	4.1 $\pm$ 0.7	3.9 $\pm$ 0.9	0.24
	Collect information from measurable patient data	3.9 $\pm$ 0.8	4.2 $\pm$ 0.7	4 $\pm$ 0.8	0.45
To process information	Analyse data to reach an understanding of signs or symptoms	4.1 $\pm$ 0.7	4.0 $\pm$ 0.7	3.9 $\pm$ 0.7	0.69
	Distinguish between relevant and irrelevant information	3.7 $\pm$ 0.8	3.7 $\pm$ 0.7	3.8 $\pm$ 0.8	0.96
To identify problems/issues	Make nursing diagnoses	3.8 $\pm$ 0.8	4 $\pm$ 0.8	3.7 $\pm$ 0.8	0.61
	Take decisions on patient care independently	3.4 $\pm$ 0.9	3.9 $\pm$ 0.6	3.7 $\pm$ 0.9	0.17
	Take decisions on patient care in cooperation with other students	2.6 $\pm$ 1.0	3.3 $\pm$ 1.0	3.5 $\pm$ 0.9	0.01 [£]
	Take decisions on patient care promptly	3.5 $\pm$ 0.8	3.72 $\pm$ 0.8	3.9 $\pm$ 0.9	0.27
To establish goals	Prioritize patient's needs for care	3.4 $\pm$ 1.1	3.9 $\pm$ 0.7	3.7 $\pm$ 1.0	0.21
	Set goals	3.3 $\pm$ 0.8	3.9 $\pm$ 0.6	3.5 $\pm$ 1.0	0.02 [£]
	Plan nursing interventions	3.7 $\pm$ 0.9	3.9 $\pm$ 0.7	3.5 $\pm$ 0.9	0.23
To take action	Implement nursing interventions	3.8 $\pm$ 1.0	4.0 $\pm$ 0.8	3.5 $\pm$ 1.1	0.35
To evaluate outcomes	Evaluate the effectiveness of interventions	3.5 $\pm$ 1.0	3.6 $\pm$ 0.8	3.4 $\pm$ 1.1	0.69
Total score (/70)		50.8 $\pm$ 6.4	54.8 $\pm$ 6.4	52.2 $\pm$ 9.5	0.17

Table 4: Scores and playing time for the three groups.

Results are described by means $\pm$ SD and compared using ANOVA (followed by post hoc comparison when significant).

*: p value < 0.05 was considered significant between Groups S, R and E.

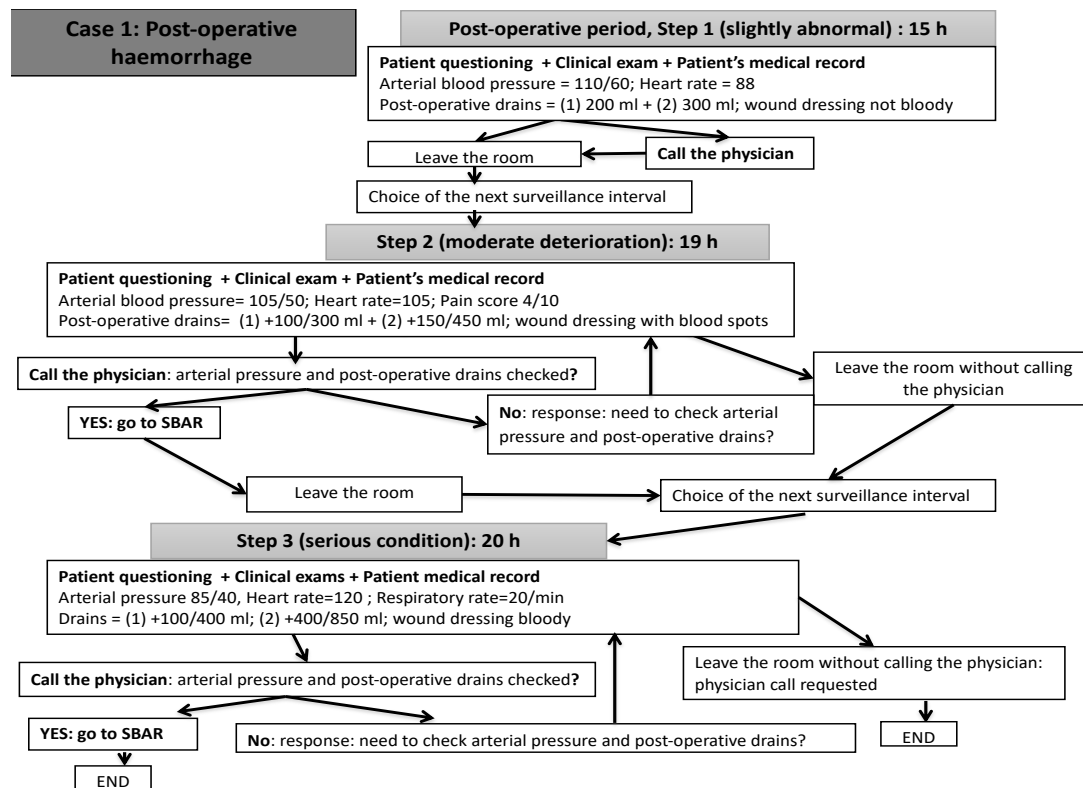
[£]: n = 24 because one student experienced a technical problem with data recording in Case 1 which could not be stored for analysis.

The participant's actions were assigned neutral, negative or positive points, as defined by the pedagogical team. The score and playing time were generated automatically by the serious game software.

	Group S n=25	Group R n=25	Group E n=21	p
Scores (score: 50 to 100)				
Post-operative haemorrhage scenario	45.5 $\pm$ 24.7 [£]	48.0 $\pm$ 19.6	41.9 $\pm$ 28.9	0.69
Brain trauma scenario	31.9 $\pm$ 31.3	38.4 $\pm$ 38	39.1 $\pm$ 44.5	0.77
Obstructed intestinal tract scenario	42.6 $\pm$ 21.6	33.7 $\pm$ 22.9	45.7 $\pm$ 21.5	0.16
Playtime (minutes)				
Post-operative haemorrhage scenario	25.7 $\pm$ 6.3 [£]	24.9 $\pm$ 7.3	25.6 $\pm$ 9.6	0.92
Brain trauma scenario	33.5 $\pm$ 7.7	28.5 $\pm$ 6.5	31.7 $\pm$ 10.5	0.10
Obstructed intestinal tract scenario	30 $\pm$ 7.3	29.4 $\pm$ 9.0	28.3 $\pm$ 7.4	0.75

Appendix 1 Figure 6 de la thèse (p 40)

Appendix 2



Scoring grid (Total score: .../100)

Detailed score at step 1, 2 and 3

Detailed score at step 1 (mildly abnormal) Items according to the participant action	Points before calling the physician	Points after calling the physician
15 h - Patient's identity checked	+5	
15 h - Pain score checked	+4	
15 h - Intravenous catheter checked	+3	
15 h - Wound dressing checked	+5	
15 h - Arterial blood pressure and heart rate checked	+5	+3
15 h – Post-operative drain 1 checked	+4	+2
15 h – Post-operative drain 2 checked	+4	+2
15 h - Non-compliance with medical prescription	-10	
15 h - Next monitoring interval: right answer	+5	
15 h - Next monitoring interval: wrong answer	-10	
TOTAL of positive points (step 1)	+35	

Detailed score at step 2 (moderate deterioration) Items according to the participant's action	Points before calling the physician	Points after calling the physician
19h- Pain score checked	+4	
19h- Wound dressing checked	+5	
19h- Wrong drug administration	-10	
19h- Arterial blood pressure and heart rate checked	+5	+3
19h- Post operative drain 1 checked	+3	+2
19h- Post operative drain 2 checked	+3	+2
19h- If physician called before checking postoperative drains or arterial blood pressure/heart rate	- 20	
19h- If physician called after checking postoperative drains and arterial pressure/heart rate	+5	
19h- 4 correct statements/questions when communicating with the surgeon in the right order (SBAR tool used)	+10	
19h- 4 correct statements/questions when communicating with the surgeon but in a wrong order (SBAR tool used)	-5	
19h- Wrong statements/questions when communicating with the surgeon and not using the SBAR tool	- 10	
19h- Leaving the room without calling the physician	-10	
19h- Next monitoring interval: right answer	+5	
19h- Next monitoring interval: wrong answer	-10	
TOTAL of positive points (step 2)	+ 40	

Detailed score at step 3 (serious condition) Items according to the participant's action	Points before calling the physician	Points after calling the physician
20 h - Wound dressing checked	+4	
20 h - Arterial blood pressure and heart rate checked	+5	+3
20 h - Post-operative drain 1 checked	+3	+2
20 h - Post-operative drain 2 checked	+3	+2
20 h - Wrong drug administration	-10	
20 h - Leaving the room without calling the physician	-30	
20 h - If physician called before checking post-operative drains or arterial blood pressure/heart rate.	-30	
20 h - If physician called after checking post-operative drains or arterial blood pressure/heart rate.	+5	
20 h - Four correct statements/questions when communicating with the surgeon in the correct order (SBAR tool used)	+5	
20 h - Four correct statements/questions when communicating with the surgeon but in the incorrect order (SBAR tool used)	-5	
20 h - Incorrect statements/questions when communicating with the surgeon and failure to use the SBAR tool	-10	
TOTAL of positive points (step 3)	+25	

Appendix 3

a) Factorial analysis (Principal Component Analysis, PCA).

(a) for scores and players' experience

(b) for scores and questions about content validity and face validity (see scores and questions in Tables 2 and 3).

* indicates loadings > 0.7 (i.e. the factor explains >49% of the item variance)

Commonalities estimate the proportion of variance of an item explained by all factors.

(a)

	Factor 1	Factor 2	Commonalities h ²
Experience with post-operative haemorrhage	0.958*	0.062	0.922
Experience with brain trauma	0.938*	0.038	0.881
Experience with intestinal obstruction	0.907*	-0.053	0.825
Score for post-operative haemorrhage scenario	-0.105	0.742*	0.562
Score for brain trauma scenario	0.043	0.821*	0.676
Score for obstructed intestinal tract scenario	0.097	0.728*	0.539
Explained variance	44.1%	29.4%	73.5%

(b)

Scores and questions of Tables 2 and 3	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Communalities h ²
Q1 (post-operative haemorrhage scenario)	0.711*	0.078	0.190	0.247	-0.101	0.619
Q1 (brain trauma scenario)	0.806*	0.222	0.056	0.059	0.224	0.756
Q1 (obstructed intestinal tract scenario)	0.821*	0.139	0.046	-0.016	-0.109	0.708
Q2	0.155	0.468	0.445	-0.186	0.234	0.530
Q3 (post-operative haemorrhage scenario)	0.144	0.130	0.804*	0.151	0.116	0.720
Q3 (brain trauma scenario)	0.166	0.023	0.810*	0.157	0.110	0.721
Q3 (obstructed intestinal tract scenario)	0.194	0.011	0.792*	0.058	0.086	0.676
Q4	0.632	0.186	0.122	0.178	-0.156	0.505
Q5 (post-operative haemorrhage scenario)	0.579	-0.004	0.593	0.208	0.008	0.730
Q5 (brain trauma scenario)	0.720*	-0.080	0.380	0.104	0.116	0.693

Q5 (obstructed intestinal tract scenario)	0.686	0.123	0.545	0.045	0.094	0.794
Q6	0.250	0.319	0.114	0.693	-0.048	0.660
Q7	0.177	0.222	0.324	0.656	-0.154	0.640
Q8	0.148	0.565	0.154	0.340	0.358	0.609
Q9	0.153	0.727*	0.087	0.342	0.052	0.679
Q10	0.064	0.838*	-0.039	0.245	-0.063	0.772
Q11	0.008	0.806*	-0.049	0.121	-0.185	0.701
Q12	0.272	0.640	0.197	0.045	0.252	0.588
Q (easy to use the game)	0.262	0.233	0.117	0.713*	0.226	0.696
Score for post-operative haemorrhage scenario	-0.186	0.104	-0.055	0.627	0.539	0.732
Score for brain trauma scenario	-0.078	-0.060	0.100	0.014	0.845*	0.734
Score for obstructed intestinal tract scenario	0.082	0.132	0.409	0.057	0.611	0.568
Explained variance	18.2%	14.6%	15.3%	10.7%	8.6%	67.4%

b) Factorial analysis (Principal Component Analysis, PCA) testing various aspects of the questionnaire (realism, educational content and impact on the player) (scores and questions in Tables 2 and 3).

* indicates loadings > 0.7 (i.e., the factor explains >49% of the item variance)

Communalities estimate the proportion of variance of an item explained by all factors.

	Factor 1	Factor 2	Factor 3	Factor 4	Communalities h ²
Q1 (post-operative haemorrhage scenario)	0.734*	0.093	0.176	0.197	0.617
Q1 (brain trauma scenario)	0.763*	0.253	0.147	0.087	0.675
Q1 (obstructed intestinal tract scenario)	0.827*	0.134	0.040	0.014	0.704
Q2	0.103	0.428	0.489	-0.029	0.434
Q3 (post-operative haemorrhage scenario)	0.144	0.141	0.824*	0.120	0.734
Q3 (brain trauma scenario)	0.149	-0.004	0.795*	0.235	0.709
Q3 (obstructed intestinal tract scenario)	0.178	0.005	0.800*	0.086	0.679
Q4	0.649	0.118	0.031	0.285	0.517
Q5 (post-operative haemorrhage scenario)	0.575	-0.014	0.576	0.246	0.723
Q5 (brain trauma scenario)	0.710*	-0.069	0.401	0.136	0.688

Q5 (obstructed intestinal tract scenario)	0.658	0.125	0.575	0.099	0.789
Q6	0.214	0.288	0.073	0.751*	0.698
Q7	0.183	0.163	0.198	0.738*	0.644
Q8	0.064	0.588	0.270	0.363	0.555
Q9	0.122	0.734*	0.100	0.354	0.689
Q10	0.071	0.835*	-0.066	0.232	0.760
Q11	0.060	0.790*	-0.122	0.098	0.652
Q12	0.243	0.667	0.279	0.047	0.584
Q (easy to use the game)	0.169	0.243	0.160	0.756*	0.685
Explained variance	19.9%	16.6%	17.5%	12.1%	66.1%

c) Factorial analysis (Principal Component Analysis, PCA) between the clinical reasoning process components (self-questionnaire) and errors made during the game (communication and situational awareness). Two students were excluded (one because data on the post-operative haemorrhage scenario were lacking and another because he/she did not answer the questionnaire).

* indicates loadings > 0.7 (i.e. the factor explains >49% of the item variance)

Communalities estimate the proportion of variance of an item explained by all factors.

Global population

	Factor 1	Factor 2	Factor 3	Communalities h²
Situational awareness errors	-0.012	0.010	-.948*	0.899
Communication errors	0.009	0.950*	-0.012	0.903
Collect	0.762*	0.272	0.006	0.655
Process	0.665	0.054	0.187	0.480
Identify	0.785*	-0.145	0.151	0.660
Establish goal	0.818*	0.057	-0.181	0.705
Take action	0.700*	-0.248	-0.207	0.594
Explained variance	40.0%	15.2%	14.8%	70.0%

Student nurse (S) group

	Factor 1	Factor 2	Communalities h²
Situational awareness errors	-0.058	-0.447	0.203
Communication errors	0.487	-0.309	0.333
Collect	0.287	0.735*	0.623
Process	-0.137	0.834*	0.714
Identify	0.568	0.418	0.497
Establish goal	0.834*	0.247	0.757

Take action	0.773*	0.027	0.598
-------------	--------	-------	-------

Explained variance	28.0%	25.2%	53.2%
--------------------	-------	-------	-------

Recently graduated (R) group

	Factor 1	Factor 2	Factor 3	Communalities h ²
Situational awareness errors	0.071	0.025	0.963*	0.933
Communication errors	-0.693	0.365	-0.091	0.622
Collect	0.195	0.868*	0.122	0.806
Process	0.094	0.898*	-0.099	0.825
Identify	0.666	0.333	-0.315	0.654
Establish goal	0.807*	0.282	0.059	0.734
Take action	0.745*	0.181	0.118	0.602
Explained variance	31.2%	27.4%	15.4%	74.0%

Expert nurse (E) group

	Factor 1	Factor 2	Factor 3	Communalities h ²
Situational awareness errors	0.076	-0.021	0.992*	0.990
Communication errors	-0.042	-0.943*	0.017	0.891
Collect	0.750*	-0.417	0.036	0.738
Process	0.856*	0.213	0.128	0.794
Identify	0.899*	0.128	-0.144	0.845
Establish goal	0.853*	-0.256	0.081	0.800
Take action	0.816*	0.064	0.152	0.693
Explained variance	50.1%	17.1%	15.0%	82.2%

3.3. Conclusion

Cette 1ère étude a évalué la validité du jeu sérieux *LabForGames Warning* selon le cadre théorique proposé par Messick. Cette étude a démontré que les scores et le temps de jeu de *LabForGames Warning* ne pouvaient pas différencier le niveau des compétences cliniques des infirmiers. Cependant, les preuves de validité ont été obtenues pour le contenu, le processus de réponse et la structure interne. Même si cette version du jeu ne peut donc être utilisée pour une évaluation sommative des étudiants infirmiers, notre étude montre que ce jeu sérieux est bien accepté par les étudiants et qu'il peut être utilisé pour la formation au sein d'un programme éducatif. D'autres études seront nécessaires pour améliorer le détail du score du jeu et plus généralement définir la place de ce nouvel outil dans le champ de l'enseignement.

4. Étude 2 : Évaluation de l'efficacité du jeu sérieux sur l'apprentissage du raisonnement clinique nécessaire à la détection de l'aggravation d'un patient chez les étudiants infirmiers en comparant une simulation avec jeu sérieux et un enseignement traditionnel.

4.1. Introduction : raisonnement clinique et détérioration de l'aggravation clinique d'un patient

La détection de l'aggravation clinique d'un patient et la communication interprofessionnelle sont des enjeux majeurs de santé publique^{99,103,118,119}. Il est donc indispensable de former les infirmiers, et pour cette raison des stratégies éducatives ont été développées^{97,120,121}. L'équipe de Liaw et al. a développé tout un programme de formation : *Rescuing A Patient In Deteriorating Situations (RAPIDS)*, programme associant la théorie à la simulation (simulation haute-fidélité, jeu sérieux...) ^{34,93,122}. De même, l'équipe Australienne a développé un programme *FIRSTACT WEB* avec le même type d'outils pédagogiques^{86,97}. Jusqu'à présent en France, cette thématique n'était abordée que dans le cadre d'un enseignement traditionnel (avec des cours) au sein des IFSI. Suite à l'essor de la simulation, des équipes ont commencé à réaliser des sessions de simulation haute-fidélité. L'équipe de LabForSIMS en partenariat avec les équipes de quatre IFSI a monté le projet de jeu sérieux *LabForGames Warning* ciblant cette thématique. Cependant les modalités pédagogiques devaient être mieux définies. Dans notre étude précédente, nous avons montré que ce jeu sérieux était bien accepté par les étudiants et qu'il pouvait être utilisé pour la formation au sein d'un programme éducatif. Le raisonnement clinique est particulièrement important dans la détection et à la réponse efficace devant l'aggravation d'un patient. Cooper et al. rapportent un déficit significatif dans le domaine du raisonnement clinique et de la capacité à détecter et prendre en charge la détérioration d'un patient chez les étudiants infirmiers¹²³. L'objectif de cette 2^{ème} étude a donc été d'évaluer l'efficacité de la simulation numérique avec *LabForGames Warning* sur l'apprentissage du raisonnement clinique en comparant un groupe d'étudiants infirmiers formé par le jeu sérieux par rapport à un groupe formé par un enseignement traditionnel.

4.2. Article 2: 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 students¹²⁴.

RESEARCH ARTICLE

Open Access



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

Antonia Blanié^{1,2,3,4*} , Michel-Ange Amorim^{3,4} and Dan Benhamou^{1,2,3,4}

Abstract

Background: Early detection and response to patient deterioration influence patient prognosis. Nursing education is therefore essential. The objective of this randomized controlled trial was to compare the respective educational value of simulation by gaming (SG) and a traditional teaching (TT) method to improve clinical reasoning (CR) skills necessary to detect patient deterioration.

Methods: In a prospective multicenter study, and after consent, 2nd year nursing students were randomized into two groups:

- Simulation by gaming “SG”: the student played individually with a serious game consisting of 2 cases followed by a common debriefing with an instructor;
- Traditional Teaching “TT”: the student worked on the same cases in text paper format followed by a traditional teaching course with a PowerPoint presentation by an instructor.

CR skill was measured by script concordance tests (80 SCTs, score 0–100) immediately after the session (primary outcome) and on month later. Other outcomes included students’ satisfaction, motivation and professional impact.

Results: One hundred forty-six students were randomized. Immediately after training, the SCTs scores were 59 ± 9 in SG group ($n = 73$) and 58 ± 8 in TT group ($n = 73$) ($p = 0.43$). One month later, the SCTs scores were 59 ± 10 in SG group ($n = 65$) and 58 ± 8 in TT group ($n = 54$) ($p = 0.77$). Global satisfaction and motivation were highly valued in both groups although significantly greater in the SG group ($p < 0.05$). The students declared that the training course would have a positive professional impact, with no difference between groups.

Conclusions: In this study assessing nursing student CR to detect patient deterioration, no significant educational difference (SCT), neither immediate nor 1 month later, was observed between training by SG and the TT course. However, satisfaction and motivation were found to be greater with the use of SG.

Trial registration: ClinicalTrials.gov; [NCT03428269](https://clinicaltrials.gov/ct2/show/study/NCT03428269). Registered 30 january 2018.

Keywords: Serious games, Simulation, Clinical reasoning, Motivation, Nursing students, Deterioration, Script concordance test

* Correspondence: antonia.blanie@aphp.fr

¹Centre de simulation LabForSIMS, Faculté de médecine Paris Saclay, 94275 Le Kremlin Bicêtre, France

²Département d’Anesthésie-Réanimation, CHU Bicêtre, 78, rue du Général Leclerc, 94275 Le Kremlin Bicêtre, France

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Patient deterioration detection is a major healthcare problem. Indeed, acute patient clinical deterioration is often preceded by a modification of physiological parameters 6 to 24 h before the event [1, 2]. The association of i) early detection ii) rapidity of response and iii) quality of clinical response, influence the patient's prognosis. Many studies have shown that delayed diagnosis of an ongoing complication increases morbidity and mortality [2, 3]. The education of nurses, who are frontline healthcare providers, is therefore essential. In French nursing institutes, theoretical training to detect patient deterioration is currently performed by traditional teaching courses and/or paper case-based courses. Simulation-based education is recommended for the training of healthcare professionals [4–6]. When compared with high-fidelity simulation, serious games also possess an interesting immersive capacity and may be used to train a large number of healthcare professionals in a limited amount of time with reduced educational resources.

Serious games are video games developed specifically with an educational purpose [7, 8]. They can be computer-based or use more immersive technologies such as virtual reality combined with head-mounted display. Serious games can promote experiential learning, as described by Kolb [9], when they include a 3D realistic environment close to real life. They can be used to train both technical and non-technical skills [10–14]. We developed a serious game, named *LabForGames Warning*, which aims to improve nursing students' inter-professional communication behavior and their capabilities to detect patient clinical deterioration. Several studies using serious games in healthcare have already been published and have been included in meta-analyses [10–14], but careful analysis shows that serious games are difficult to compare due to the differences in the populations studied, the variety of game designs, topics included, pedagogical objectives and modalities of assessment. Comparative analysis between the educational value of a serious game and a traditional course is already available for cardiac arrest management, trauma triage and other domains, but it is uncertain whether results obtained in one field apply to the others [13–18]. In addition, several studies have used a serious game to train nurses in the detection of patient deterioration, but are characterized by a high level of heterogeneity and sometimes provide neutral results [19–22]. One randomized study, aiming to compare the assessment and the management of clinical deterioration, in which either a virtual patient or a mannequin was used, found that both simulation methods (serious game and high-fidelity) could be effective in improving nursing-student score performance, with no difference between the two methods [23]. More specifically, few studies have

assessed how components of clinical reasoning skills necessary to detect patient deterioration are modified by the use of serious games. Most serious games studies use one result of clinical reasoning (treatment, diagnosis or triage) but an objective criterion evaluating all clinical reasoning skills has rarely been used so far [15, 20]. As described by Levett-Jones et al. [24], clinical reasoning is “the process by which nurses collect cues, process the information, come to an understanding of a patient problem or situation, plan and implement interventions, evaluate outcomes, and reflect on and learn from the process”. Furthermore, the positive impact of instructor-standardized debriefing is clearly demonstrated in high fidelity simulation [25] but its place is unclear regarding the use of a serious game [23, 26].

The objective of this study was to compare the respective value of simulation using the above-mentioned serious game with debriefing and a traditional teaching method to improve the clinical reasoning skills necessary to detect patient deterioration in nursing students.

Methods

Serious game development

The serious game project was promoted by the simulation center of Paris-Sud University (LabForSIMS) in collaboration with four nursing institutes (Sud Francilien, Perray Vacluse, Paul Guiraud and Etampes) through a grant from the Ile-de-France Regional Health Agency (ARS). Four virtual clinical cases were created through an iterative dialogue between the pedagogical team and the software designer (Interaction Healthcare®, Levallois-Perret, France). The pedagogical team were clinical experts (instructors of four nursing schools and anesthesiologists) who were also involved in the simulation center. The predefined educational objectives of the serious game were detection of clinical deterioration and communication. In the game, a nurse was expected to identify clinical deterioration in different clinical situations and warn the medical team appropriately, in consideration of the patient's clinical severity. As the serious game focuses on nursing students, the objectives had to comply with the French official nursing repository [27]. The game is not played in real time but a clock was presented on the screen which indicated the time flow. In each clinical scenario, three consecutive steps (mildly abnormal, moderate aggravation and severe condition) were constructed to reproduce a specific complication of increasing severity, with the aim of introducing the concept of early warning signs [28]. In the present study, only two of the four available cases were used (Additional file 1):

- Postoperative hemorrhage case: an adult female patient who had undergone a scheduled total hip

replacement earlier in the day and who is lying in her ward room bed immediately after arrival from the post-anesthesia care unit. Postoperative hemorrhage is progressively occurring from the surgical site.

- Brain trauma case: an elderly patient with dementia living in a nursing home, whose anticoagulation is associated with progressively developing neurological deterioration after brain trauma from a fall.

Learning safe and standardized communication was an additional educational objective [29, 30]. We chose to train nursing students to the SBAR method, (« Situation, Background, Assessment, Recommendation »), adapted into French by the French Health Authority [27].

Before starting to play, a tutorial allowed a first approach to the software. In the game, the player was a nurse. The nurse played on a computer and could move in the 3-D environment (Additional file 1). The student's actions were obtained by clicking on different interactive areas. For example, regarding clinical examination, the student could check arterial pressure, motor response, etc. Patient questioning was performed by a scroll-down menu of pre-determined options but the player could hear the patient's answer. The student had also access to the patient's file (prescription, nurse handover transcripts ...) and could call the physician with a virtual phone and present the case using the SBAR method.

At the end of each scenario, an automatic feedback was presented to the student, which included main guidelines and key messages as well as a global and detailed score according to a grid which had been constructed by the pedagogical team during the development phase of the SG. The student's actions regarding clinical examination (e.g., checking of arterial pressure, pain assessment...) and the student's decision (e.g., call for the physician...) were scored using positive, negative or neutral points depending on the different steps of the case. Moreover, positive or negative points were attributed to the quality of the communication during the SBAR tool part of the game. This score and the playtime were not used to evaluate the students in this present study.

Study description

After informed and written consent, 2nd year nursing students were included and randomized into two groups: the simulation by gaming group (SG group) and the traditional teaching group (TT group) (Fig. 1).

Training sessions for the detection of clinical deterioration were planned in the four nursing institutes. The group sessions lasted 2 h and each involved 15 students. The teachers were nurse instructors involved in the

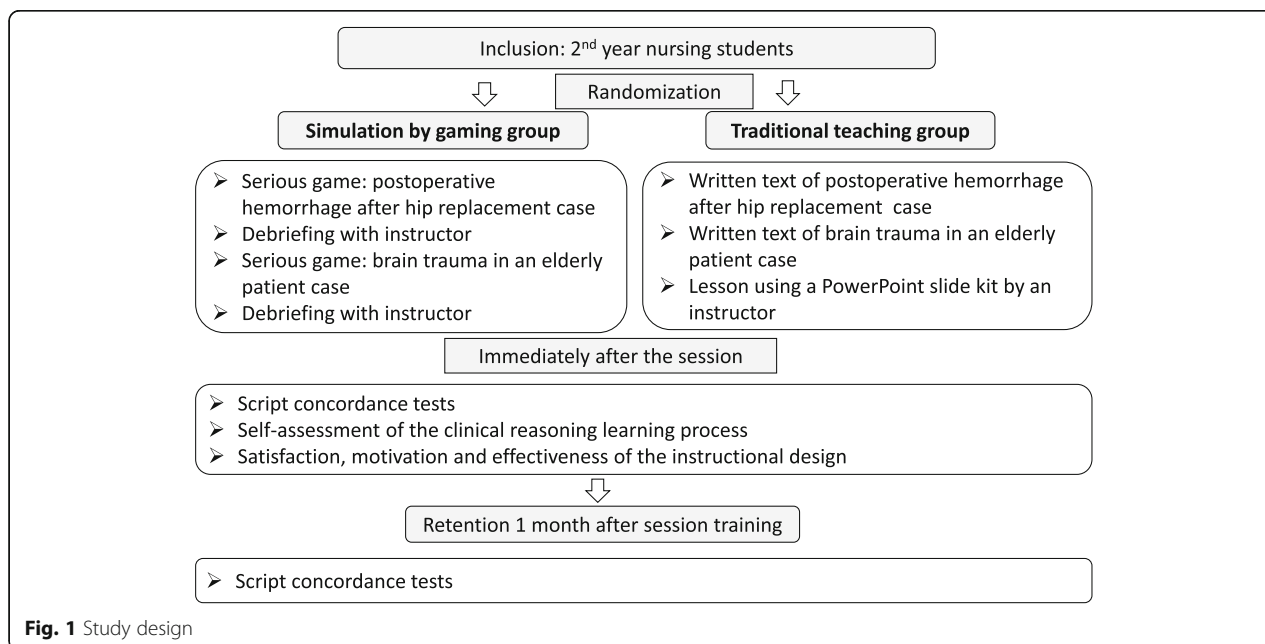
program and trained in simulation. The students had not been informed of the study objectives, nor to which group they were to be assigned (single-blind). Randomization was carried out on the morning preceding the session with the use of the random Excel function (by A.B.). The sessions and assessment for the two groups were run simultaneously in different rooms.

In the SG group, the students played individually on computers and were given the two previously described cases of *LabForGames Warning* (Additional file 1). A whole-group debriefing was performed after the cases by trained instructors. Debriefing (different from serious game's automatic feedback) was performed according to the methodology used in a simulation session (reaction, analysis and synthesis phases) [25, 31, 32] in which reflexive practice is promoted by a positive interaction between students and instructor.

In the traditional teaching group, the students individually analyzed the same 2 cases which were presented in text paper format (Additional file 2) (without instructor). Then, they attended a PowerPoint slide kit traditional teaching course on the topic with an instructor (Additional file 3).

Assessment method

All players answered a questionnaire at the end of the session which included questions on sex, age, post-graduate experience and previous video gaming activity (entertainment and professional education). The primary outcome measure was the student's clinical reasoning skills regarding detection of clinical deterioration as measured by script concordance tests (SCTs) immediately after the session. SCTs have been validated as an effective tool of assessing clinical reasoning skills (Kirkpatrick level 2) and are considered to be an objective and quantitative assessment method. SCTs have already been used in nursing students [33–37]. Based on the script theory [37], SCTs are used to compare if students' decisions, made from their knowledge networks (scripts), are in line with the decisions taken by a panel of experts. The SCTs used in this study were carefully prepared by following the guideline described by Charlin et al. and Fournier et al. to obtain a high level of fidelity and validity [33–35]. An example of SCT is presented in Table 1. SCTs are case-based tests consisting of short scenarios, and for each the trainee has to interpret newly formulated information against the baseline one to modulate the final decision with the use of a five-point Likert-scale. This final decision could be a diagnosis but also a treatment, an action (prescribe or perform a complementary exam, call for help ...). Because the numerous short cases used explore the different types of decision, the SCTs test the different skills of clinical deterioration reasoning and not just the diagnosis.



The standard SCT scoring method uses the aggregate method in which scoring is in relation to the scores obtained by a panel of reference experts in the respective domain. After scoring the single items, a raw summary score is obtained by summing up all of cases. In order to make the scores interpretable and comparable, the raw summary score is normalized where the expert mean performance is set at 80 points, and the standard deviation at 5 points. This normalization is a Z transformation as suggested by Charlin et al. [38]. The analysis of SCTs includes variance calculation and scores obtained by an aggregation method using the following file (https://www.cpass.umontreal.ca/recherche/groupe-de-recherche-cpass/axes-de-recherches/concordance/tcs/corriger_tcs/). Ninety-three SCTs in connection with the pedagogical objectives were constructed by 3 expert instructors (2 anesthesiologists and 1 nurse instructor). Eleven other experts were subsequently included in the reference panel (6 anesthesiologists and 5 nurse instructors). After analysis of the expert panel responses, 13

SCTs were removed because of high variance (variance > 1) and 80 SCTs were finally used. After optimization of the tool, Cronbach's alpha was 0.75 (95 CI: 0.70–0.81). The mean SCT score obtained by experts ($n = 11$) was 83.3 ± 3.6 (95 CI: 82.2–84.4). The students were trained to use the SCT format just before the training session. The same 80 SCTs were applied immediately and 1 month after the session, to capture retention of knowledge, with the use of an online module.

In addition, secondary outcome measures were assessed. A self-assessment of the perceived change of the clinical reasoning process was recorded after the session to assess the various steps of nursing clinical reasoning as defined by Levett-Jones et al. [24]. The questionnaire initially described by Koivisto et al. was modified by adding 2 specific questions on clinical deterioration described by Liou et al. [20, 39]. The modified questionnaire was translated into French by the research group (A.B. and D.B.). Each question assesses a specific step of the clinical reasoning process ("I learned

Table 1 Example of script concordance test

N°	If you were thinking of...	And then you find that ...	this hypothesis becomes ... (circle your answer)				
1.1	Anticipating the drawing of a blood count	His heart rate is 120 beats/ min	-2	-1	0	1	2
1.2	drawing a capillary blood glucose	His blood pressure is 80 / 40 mmHg	-2	-1	0	1	2
1.3	Calling the physician	He has abdominal pain	-2	-1	0	1	2

You are a nurse in the Surgery Department

A 45-year-old patient who has undergone splenectomy 2 days ago calls you because he feels some discomfort. He has a history of hypertension

- 2 Much less likely

- 1 Less likely

0 Neither more nor less likely

+ 1 More likely

+ 2 Much more likely

to”) using a five-point Likert scale. A global score (graded out of 75) was obtained by adding values given to the 15 questions [20, 24, 39]. Moreover, the students’ perceived satisfaction, their motivation toward learning the specific topic and the effectiveness of the instructional design were assessed by a questionnaire at the end of the session with a Likert scale (1 to 10) corresponding to level 1 and 3 of the Kirkpatrick training evaluation model [40].

Statistical analysis

The primary outcome was the students’ clinical reasoning skills regarding the detection of clinical deterioration as measured by the SCTs. We assumed that the traditional teaching group would reach an average score of 65/100 together with a standard deviation of 10/100 and, that the new training modality would improve the score by at least one standard deviation (mean difference 10/100 before-after). Accordingly, the sample size was set to 50 students per group using an alpha risk of 5% and beta risk of 10% with a two-sided two sample t-test (R software, <https://marne.u707.jussieu.fr/biostatgy/>). In view of the risk of attrition, each group was composed of 73 students.

The results are presented as mean $\pm$ standard deviation or percentage and confidence intervals. After assessing normal distribution, statistical analysis was performed with the use of parametric tests (Student’s *t* test or χ^2 test) (JMP software, SAS institute®). A *p* value less than 0.05 was considered significant. For multiple comparisons, a Bonferroni correction was applied and the statistical significance threshold was lowered to 0.003 (alpha/15).

Ethical statement

This study was approved by The Institutional Review Board of SFAR (IRB-00010254 – 2017-044). The project had been registered on ClinicalTrials.gov (NCT03428269) [41]. The study was carried out with the use of the CONSORT tool adapted for simulation studies and the GREET Tool for educational studies [42].

Results

Participants

Five training sessions were organized in the 4 nursing institutes in February 2018. In total, 146 voluntary nursing students were included and randomized: *n* = 73 in the SG group and *n* = 73 in the TT group. No exclusion was observed (Fig. 2). Participant characteristics are presented in Table 2 and no significant differences were observed. The students of both groups did not differ in their experience regarding the clinical situations presented.

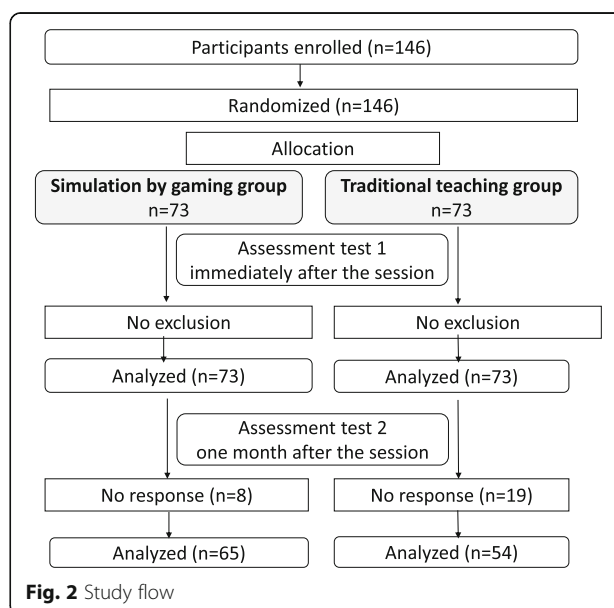


Fig. 2 Study flow

SCTs results immediately after the session (primary outcome) and 1 month later

Immediately after the session, the mean SCT scores were 58.9 ± 9.1 in the SG group and 57.8 ± 8 in the TT group with no significant difference ($p = 0.43$) (Fig. 3). One month later, 119 nurse students answered the SCTs ($n = 119/146$): $n = 65$ in the SG group and $n = 54$ in the TT group. The mean SCT scores were 58.5 ± 10.2 in the SG group and 58 ± 9.1 in the TT group ($p = 0.77$) (Fig. 3). Scores obtained immediately after the session and 1 month later were not significantly different between groups.

Self-assessment of clinical reasoning

Following the training session, all students said that their knowledge of the different steps of the clinical reasoning process had increased. The scores were all above 3.4/5 with no significant difference between groups (Table 3).

Satisfaction, motivation and professional transfer

Students in the SG group significantly expressed significantly more satisfaction toward the training session than those in the TT group ($p = 0.001$) (Table 4).

Moreover, regarding the pedagogical tool used for training, the students of the SG group expressed more satisfaction than those in the TT group ($p = 0.04$). Students of the SG group perceived the training session as more engaging, as reflected by their significantly increased motivation ($p = 0.003$).

The global educational value (*Would you recommend this training to students or colleagues?*) was more positively significant in the SG group ($p = 0.002$). Both

Table 2 Participant characteristics

	SG group (n = 73)	TT group (n = 73)	p
Age (years) (mean ± SD)	24 ± 6.4	25 ± 6.5	0.39
Sex (F/M) n (%)	59/14 (81/19%)	65/8 (89/11%)	0.17
Video gaming activity: n (%)			
Never	39 (53%)	43 (59%)	0.15
1/month	11 (15%)	12 (16%)	
1/week	10 (14%)	13 (18%)	
Everyday	13 (18%)	4 (6%)	
No response	0 (0%)	1 (1%)	
Video gaming activity in healthcare: % of players			
Yes	13 (18%)	13 (18%)	0.97
No	59 (81%)	60 (82%)	
No response	1 (1%)	0 (0%)	
Previous experience of clinical situations presented in the course: 1 (no) to 10 (expert)			
Brain trauma	2.7 ± 2.1	2.3 ± 1.8	0.21
Postoperative hemorrhage	3.0 ± 2.4	2.6 ± 2.0	0.39

Results are presented as mean ± standard deviation, or percentage. Comparisons were carried out with the use of the Student's *t* test on means, or a Chi² test on percentages

*: *p* value less than 0.05 was considered significant

groups declared that the session could have an impact on their future professional work but the difference failed to reach significance (Table 4).

Discussion

In this study in which we assessed clinical reasoning skills to detect patient deterioration, we found no significant difference between a serious game-based simulation format and a traditional teaching method immediately and 1 month after training. However, the students expressed more satisfaction and motivation with the innovative teaching method.

The education of nurses, who are frontline health-care providers, is essential to improve the detection of patient deterioration. Serious games dedicated to the same learning objective have already been created by others [19–22]. In Australia, a web-based e-simulation program suggested an improvement of clinical knowledge of patient deterioration and enhanced students' self-assessed of knowledge, skills, confidence, and competence [19]. Similar results were obtained by Liaw et al. in nurses working on surgical wards [21, 22] and another team also used a serious game to explore the learning process of nursing

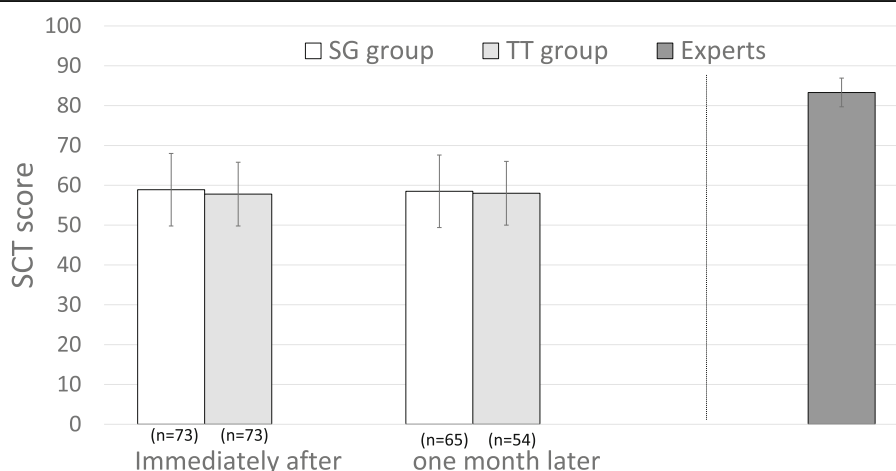


Fig. 3 SCT score immediately and 1 month after the session in SG group and TT group

Table 3 Results of self-assessment of learning the clinical reasoning process between groups

I learned: 1 (not at all) to 5 (very much)	SG group (n = 73)	TT group (n = 73)	p
To collect information			
Collect information by interviewing the patient	4.0 ± 1.0	3.6 ± 1.2	0.02
Collect information by observing the patient	4.2 ± 0.8	4.0 ± 1.0	0.19
Collect information from measurable patient data	4.2 ± 0.8	4.2 ± 0.9	0.76
To process information			
Analyze data to reach an understanding of signs or symptoms	4.3 ± 0.8	4.3 ± 0.8	0.86
To identify problems/issues			
Make nursing diagnosis	4.2 ± 0.8	4.0 ± 0.9	0.10
Recognize possible early signs or symptoms when a patient's health deteriorates	4.1 ± 0.8	4.1 ± 0.9	0.91
Make decisions on patient care independently	4.1 ± 0.8	3.7 ± 0.8	0.004
Make decisions on patient care in cooperation with other students	3.5 ± 1.2	3.0 ± 1.1	0.75
Make prompt decisions on patient care	3.9 ± 0.9	3.8 ± 0.9	0.30
To establish goals			
Prioritize patient's need for care	3.7 ± 1.0	3.9 ± 1.1	0.24
Set goals	3.4 ± 1.0	3.4 ± 1.2	0.82
Plan nursing interventions	3.5 ± 1.1	3.6 ± 1.1	0.54
To take action			
Implement nursing interventions	4.1 ± 0.9	4.1 ± 0.7	0.61
Communicate vital information clearly based on the patient's current condition	4.4 ± 0.8	4.3 ± 0.9	0.42
To evaluate outcome			
Evaluate effectiveness of interventions	3.9 ± 0.9	3.8 ± 1.1	0.44
Total scoring (/75)	59.5 ± 0.9	57.8 ± 1.0	0.25

Results are presented as mean ± standard deviation and comparisons were carried out with the use of the Student's *t* test. The Bonferroni criterion was set at $\alpha/15 = 0.003$ to reach statistical significance

students 'clinical reasoning, using a self-assessment questionnaire [20]. However, none of the above-mentioned studies included a traditional educational modality as a control group. Moreover, an objective and quantitative assessment method such as SCT has rarely been used [43].

In education, the learning process can be explored using the learning levels described by Kirkpatrick [40]. Our study showed no difference between the two

methods (simulation with serious game vs traditional teaching) with regard to the improvement of clinical reasoning skills necessary to detect patient deterioration (Kirkpatrick level 2). This result was obtained by comparing not only SCT scores, but also by some questions measuring the self-assessment of the different steps of the clinical reasoning process. Our results are consistent with those of Dankbarr et al. and show that serious games and e-modules are equally effective in developing

Table 4 Results of satisfaction, motivation and professional impact self-assessment

	SG group (n = 73)	TT group (n = 73)	p
Are you globally satisfied with this training course? 1 (not satisfied) to 10 (very satisfied)	8.5 ± 1.6	7.6 ± 1.7	0.001*
Are you globally satisfied with the educational tool used for case-based learning? 1 (not satisfied) to 10 (very satisfied)	8.5 ± 1.4	8.0 ± 1.6	0.04
Do you think that this training course motivates you to learn? 1 (absolutely not) to 10 (agree absolutely)	8.7 ± 1.9	7.7 ± 2.0	0.003*
Do you think that this training will have an impact on your future professional work? 1 (absolutely not) to 10 (agree absolutely)	8.3 ± 1.8	7.7 ± 1.7	0.06
Would you recommend this training to students or colleagues? 1 (absolutely not) to 10 (agree absolutely)	8.8 ± 1.7	7.8 ± 2.0	0.002*

Results are presented as mean ± standard deviation or percentage. Comparisons were carried out with the use of the Student's *t* test or a Chi² test. The Bonferroni criterion was set at $\alpha/5 = 0.01$ to reach statistical significance; *: *p* value less than 0.01 was considered significant

knowledge in patient safety [16]. Drummond et al. also found that training medical students in the management of cardiopulmonary resuscitation with a serious game was as effective as using an online course [17]. Likewise, in a randomized study, Liaw et al. failed to show any difference between virtual patient and mannequin-based simulation in a refresher training course on managing deterioration [23]. By contrast, the randomized study of Mohan et al. evidenced an improved decision-making capacity in trauma triage with the use of an educational video game as compared to app-based training [15]. Other studies have compared the efficacy of different educational methods but in other healthcare domains (such as trauma, urology, surgery, pharmacy) and have also provided conflicting results for learning [15, 18, 44, 45]. In a recent systematic review, serious games used in health professions seem at least effective as a learning control method and may be more effective in improving knowledge and skills [13]. One major limitation of this review was the high level of heterogeneity. The existing studies were difficult to compare and generalizability difficult to reach due to the differences in the populations studied, the variety of the serious game assessment methods, the game designs, and the pedagogical objectives [10–14]. The definition itself of a serious game is sometimes ambiguous between e-learning, virtual patient and serious game [46]. In our study, serious games are defined as video games developed specifically with an educational purpose [7]. Finally, in the 30 randomized clinical trials included in this systematic review, no study assessing the detection of clinical deterioration was included [13]. This is therefore important to better specify the place of the serious games in professional healthcare education and their educative impact, especially on professional practice.

In the studies described above, the clinical reasoning skills were poorly assessed. In most studies, the knowledge or the end-result of the clinical reasoning process (diagnostic or treatment), but not the clinical reasoning process itself, were assessed. In our study, we tried to explore the clinical reasoning skills and all different steps of the clinical reasoning process using firstly a method, i.e. SCT, which has been validated as an effective tool of assessing clinical reasoning skills [33–36, 43]. The SCTs used in our study were related to the content and the educational objectives (clinical deterioration) of the serious games used by students. This was a concern in order to link the strategy to a theoretical learning model of clinical reasoning, the script theory. The final decisions tested with our SCTs explored many different nurse decisions (diagnosis but also treatment or action to perform). SCT is an objective and quantitative assessment method which reduces interpretation bias. Secondly, self-assessment of the different steps of clinical

reasoning was obtained with the use of a modified tool [20, 39]. This tool explores the process by which nurses collect information, process the information, identify problems/issues, establish goals, take action, and evaluate outcomes. Recently, and after the beginning of our study, Liaw et al. published a study describing another specific tool used to assess clinical reasoning in clinical deterioration which, on the whole, used the same criteria as our own study [47]. The weaknesses in the previous literature as well as the neutral results (including ours) could be explained by the difficulty to assess clinical reasoning, the complexity of education, and the need to explore results in the long-term.

Learning retention is essential but has been less studied. Similar to our results, Drummond et al. found no difference in the management of cardiopulmonary resuscitation (by means of a mastery learning checklist) 4 months after training of medical students with a SG or an online course [17]. By contrast, the previously mentioned study of Mohan et al. showed an improved decision-making capacity in trauma triage 6 months after initial training with the use of an educational video game compared to traditional apps [15].

As described by Kirkpatrick, satisfaction is the first step of the learning process [40]. In the present study, the students were globally satisfied with the training received and the educational tool used. However, satisfaction was greater in the SG group. Satisfaction associated with the use of a serious game is often described but not always by using a blind comparison design [48, 49]. Boeker et al. reported significantly higher satisfaction in an urology adventure game group compared to a group using a written script [45]. Serious gaming may improve satisfaction compared to traditional learning and other modalities of digital education [13]. Moreover, our students perceived this training session as more engaging and providing significantly higher motivation in the SG group. These results collectively confirm that learners are very motivated to use serious games as they are more engaging, interactive and provide more continuous feedback than traditional learning methods [16, 45, 50, 51] or e-modules [16]. Although motivation and satisfaction are complex psychological processes, the motivational effect is important in education and might be associated with better learning outcomes [8, 52, 53]. Increased satisfaction and motivation when a new pedagogical tool is used could engage the student to learn more, and this could have long-term impact. Active learning is known to increase students' performance in various scientific domains [54] and this could also be true for simulation by gaming in the medical field.

Professional attitudes have been less studied, and there is limited evidence of a beneficial effect of using a serious game [13]. In the present study, both the SG and

the TT group similarly declared that the session could have an impact on their future professional work but no significant difference was found between the two groups (level 3 of the Kirkpatrick training evaluation model).

The present study was original as it included instructor-standardized debriefing after the serious game session in a way similar to debriefing carried out in high fidelity simulation [25]. Our objective was to evaluate a simulation method using serious games and not just the game itself. The importance of instructor-led debriefing is well demonstrated in simulation in medical education [31, 32]. However, paradoxically, its place is extremely limited in the use of serious games. This could be explained by the frequent online use of these games, which is more compatible with digital feedback [16, 19, 26]. In many studies, games are played in the presence of instructors who are able to interact with the players but surprisingly no standardized debriefing is performed [20, 43]. It is therefore unknown if the addition of debriefing to the SG session is of any educational value or not. In our study, digital feedback was included in the serious game, but an instructor-standardized debriefing with the players (reaction, analysis and synthesis phases) was added. Importantly, this did not improve the learning efficacy as results were similar to those obtained after the traditional training method. Additional studies are necessary to seek the efficacy of serious game including or not a debriefing session on clinical reasoning skills.

Limitations

One limitation in studies such as ours is the difficulty in assessing clinical reasoning skills as it is a complex process. We used SCTs as, in education literature, they are reported to be a well-validated evaluation method [33–36]. Our SCTs were carefully prepared and followed the guidelines described by Charlin et al. and Fournier et al. to obtain adequate fidelity and validity [33–36]. Although our students were trained to use SCTs before the beginning of the session, the training lasted 5 min only and it may be that some students had not fully understood how to correctly answer the questions. Nevertheless, the mean student SCT score (57.8 to 58.9) and the mean expert SCT score (83.3) in our study are similar to those previously described in the literature, which suggests that our use of SCTs was correct [43]. In addition, self-assessment was equally used to assess the various steps of the nursing clinical reasoning process but this part is less validated, as it rests on student's perception only [20, 24, 39, 47].

An additional limitation may be related to the use of only 2 cases which might have reduced the impact of the training course. However, the SCTs did explore the same theme of clinical deterioration detection in both groups. Thus, the lack of difference between the two groups

could be explained by the short duration of training (total: 2 h). In addition, no pretest knowledge was assessed as we expected a learning bias related to the use of the same SCTs. Moreover, the session could not be prolonged for practical reasons, and adding 80 SCTs would have reduced the time dedicated to the training itself. The pretest was not essential to compare the groups before, as both groups of students had the same low level of clinical experience and had the same basic knowledge (2nd year and same curriculum).

Conclusion

In this study, clinical reasoning was assessed in nursing students immediately and 1 month after a training course dedicated to the detection of patient deterioration. No significant educational difference between training with a serious game-based simulation course and a traditional teaching course was found. However, satisfaction and motivation were greater when the training involved a serious game-based simulation. Simulation by gaming could positively impact the long-term results of nurses' education on clinical reasoning.

Additional files

Additional file 1 Screenshot of *LabForGames Warning* serious game.

Additional file 2. Text paper format of the 2 cases (traditional teaching group).

Additional file 3. PowerPoint slide kit of traditional teaching course (traditional teaching group).

Abbreviations

CR: Clinical reasoning; SBAR: Situation, background, assessment, recommendation; SCT: Script concordance tests; SG: Simulation by gaming; TT: Traditional teaching

Acknowledgements

We would like to acknowledge the *LabForGames Warning* project team members (Philippe Rouleau, Bonga Barcello De Carvalho, Hélène Bertrand, Alain Cazin, Véronique Mahon, Lionel Henriques, Nathalie Léon, Vincent Lebreton, Laure Legoff, Aurélie Woda, Bertrand Bech, Alexandre Renault) and the instructors of the four nursing institutes (Sud Francilien, Perray Vaucluse, Paul Guiraud and Etampes).

Authors' contributions

AB designed the study, performed the inclusion, analyzed and interpreted the data regarding the serious game and was a major contributor in writing the manuscript. MAA analyzed and interpreted the data regarding the serious game and was a contributor in writing the manuscript. DB designed the study, analyzed and interpreted the data regarding the serious game and was a contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

Agence Régionale de Santé Ile de France (French regional authority for health regulation) which selected the serious game project and provided a grant to create the serious game project and mainly pay for the development of the game in 2014 (before this research). Agence Régionale de Santé Ile de France did not provide a grant for the research project. None of the authors received any personal financial support.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by The Institutional Review Board of SFAR (French Society of Anesthesia and Intensive Care Medicine) (IRB-00010254-2017-044). All participants were informed and consented.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre de simulation LabForSIMS, Faculté de médecine Paris Saclay, 94275 Le Kremlin Bicêtre, France. ²Département d'Anesthésie-Réanimation, CHU Bicêtre, 78, rue du Général Leclerc, 94275 Le Kremlin Bicêtre, France. ³CIAMS, Univ. Paris-Sud, Université Paris-Saclay, 91405 Orsay Cedex, France. ⁴CIAMS, Université d'Orléans, 45067 Orléans, France.

Received: 11 October 2019 Accepted: 16 January 2020

Published online: 19 February 2020

References

- Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation*. 2004;62(2):137–41.
- Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, Bishop GF, Simmons G. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med*. 2002;28(11):1629–34.
- Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg*. 2009;250(6):1029–34.
- Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978–88.
- McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. 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*. 2011;86(6):706–11.
- Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872–83.
- Bergeron BP. Developing serious games: Charles River media; 2006.
- Hamari J, Shernoff DJ, Rowe E, Collier B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. *Comput Hum Behav*. 2016;54:170–9.
- Kolb AY, Kolb DA. Learning styles and learning spaces: enhancing experiential learning in higher education. *Acad Manag Learn Educ*. 2005;4:193–212.
- Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg*. 2012;99(10):1322–30.
- Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med*. 2010;85(10):1589–602.
- Wang R, DeMaria S Jr, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. *Simul Healthc*. 2016;11(1):41–51.
- Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, Dauwels-Okutsu S, Nikolaou CK, Zary N, Campbell J, et al. Serious gaming and Gamification education in health professions: systematic review. *J Med Internet Res*. 2019;21(3):e12994.
- Gorbanev I, Agudelo-Londono S, Gonzalez RA, Cortes A, Pomares A, Delgadillo V, Yepes FJ, Munoz O. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online*. 2018;23(1):1438718.
- Mohan D, Farris C, Fischhoff B, Rosengart MR, Angus DC, Yealy DM, Wallace DJ, Barnato AE. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ*. 2017;359:j5416.
- Dankbaar ME, Richters O, Kalkman CJ, Prins G, Ten Cate OT, van Merriënboer JJ, Schuit SC. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. *BMC Med Educ*. 2017;17(1):30.
- Drummond D, Delval P, Abdenouri S, Truchot J, Ceccaldi PF, Plaisance P, Hadchouel A, Tesniere A. Serious game versus online course for pretraining medical students before a simulation-based mastery learning course on cardiopulmonary resuscitation: a randomised controlled study. *Eur J Anaesthesiol*. 2017;34(12):836–44.
- Berger J, Bawab N, De Mooij J, Sutter Widmer D, Szilas N, De Vriese C, Bugnon O. An open randomized controlled study comparing an online text-based scenario and a serious game by Belgian and Swiss pharmacy students. *Curr Pharm Teach Learn*. 2018;10(3):267–76.
- Bogossian FE, Cooper SJ, Cant R, Porter J, Forbes H, Team FAR. A trial of e-simulation of sudden patient deterioration (FIRST2ACT WEB) on student learning. *Nurse Educ Today*. 2015;35(10):e36–42.
- Koivisto JM, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: a cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. *Nurse Educ Today*. 2016;45:22–8.
- Liaw SY, Wong LF, Lim EY, Ang SB, Mujumdar S, Ho JT, Mordiffi SZ, Ang EN. Effectiveness of a web-based simulation in improving Nurses' workplace practice with deteriorating Ward patients: a pre- and Postintervention study. *J Med Internet Res*. 2016;18(2):e37.
- Liaw SY, Chng DYJ, Wong LF, Ho JTY, Mordiffi SZ, Cooper S, Chua WL, Ang ENK. The impact of a web-based educational program on the recognition and management of deteriorating patients. *J Clin Nurs*. 2017;26(23–24):4848–56.
- Liaw SY, Chan SW, Chen FG, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. *J Med Internet Res*. 2014;16(9):e214.
- Levett-Jones T, Hoffman K, Dempsey J, Jeong SY, Noble D, Norton CA, Roche J, Hickey N. The 'five rights' of clinical reasoning: an educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. *Nurse Educ Today*. 2010;30(6):515–20.
- Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc*. 2007;2(2):115–25.
- Middeke A, Anders S, Schuelper M, Raupach T, Schuelper N. Training of clinical reasoning with a serious game versus small-group problem-based learning: a prospective study. *PLoS One*. 2018;13(9):e0203851.
- Site legifrance.gouv. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020961044>
- Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, Esmonde L, Goldhill DR, Parry GJ, Rashidian A, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*. 2007;33(4):667–79.
- Nagpal K, Arora S, Vats A, Wong HW, Sevdalis N, Vincent C, Moorthy K. Failures in communication and information transfer across the surgical care pathway: interview study. *BMJ Qual Saf*. 2012;21(10):843–9.
- Mackintosh N, Sandall J. Overcoming gendered and professional hierarchies in order to facilitate escalation of care in emergency situations: the role of standardised communication protocols. *Soc Sci Med*. 2010;71(9):1683–6.
- Guide bonnes pratiques simulation sante. https://www.has-sante.fr/upload/docs/application/pdf/2013-01/guide_bonnes_pratiques_simulation_sante_guide.pdf. Accessed 28 Jan 2020.
- Salas E, Klein C, King H, Salisbury M, Augenstein JS, Birnbach DJ, Robinson DW, Upshaw C. Debriefing medical teams: 12 evidence-based best practices and tips. *Jt Comm J Qual Patient Saf*. 2008;34(9):518–27.
- Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ*. 2012;46(6):552–63.
- Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak*. 2008;8:18.
- Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CP. Script concordance testing: a review of published validity evidence. *Med Educ*. 2011;45(4):329–38.
- Deschenes MF, Charlin B, Gagnon R, Goudreau J. Use of a script concordance test to assess development of clinical reasoning in nursing students. *J Nurs Educ*. 2011;50(7):381–7.

37. Deschênes MF, Goudreau J. Addressing the development of both knowledge and clinical reasoning in nursing through the perspective of script concordance: an integrative literature review. *J Nurs Educ*. 2017;7(12):28–38.
38. Charlin B, Gagnon R, Lubarsky S, Lambert C, Meterissian S, Chalk C, Goudreau J, van der Vleuten C. Assessment in the context of uncertainty using the script concordance test: more meaning for scores. *Teach Learn Med*. 2010;22(3):180–6.
39. Liou SR, Liu HC, Tsai HM, Tsai YH, Lin YC, Chang CH, Cheng CY. The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. *J Adv Nurs*. 2016;72(3):707–17.
40. Rouse DN. Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspect Health Inf Manag*. 2011;8:1c.
41. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, Duval-Arnould J, Lin Y, Cook DA, Pusic M, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Adv Simul (Lond)*. 2016;1:25.
42. Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Moher D, Tilson JK, Williams MT. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). *BMC Med Educ*. 2016;16(1):237.
43. Schubach F, Goos M, Fabry G, Vach W, Boeker M. Virtual patients in the acquisition of clinical reasoning skills: does presentation mode matter? A quasi-randomized controlled trial. *BMC Med Educ*. 2017;17(1):165.
44. Graafland M, Bemelman WA, Schijven MP. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. *Surg Endosc*. 2017;31(10):4093–101.
45. Boeker M, Andel P, Vach W, Frankenschmidt A. Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. *PLoS One*. 2013;8(12):e82328.
46. Cant R, Cooper S, Sussex R, Bogossian F. What's in a name? Clarifying the nomenclature of virtual simulation. *Clin Simul Nurs*. 2019;27:26–30.
47. Liaw SY, Rashasegaran A, Wong LF, Deneen CC, Cooper S, Levett-Jones T, Goh HS, Ignacio J. Development and psychometric testing of a clinical reasoning evaluation simulation tool (CREST) for assessing nursing students' abilities to recognize and respond to clinical deterioration. *Nurse Educ Today*. 2018;62:74–9.
48. Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebo A. A serious game for teaching nursing students clinical reasoning and decision-making skills. *Stud Health Technol Inform*. 2016;225:905–6.
49. Pasquier P, Merat S, Malgras B, Petit L, Queran X, Bay C, Boutonnet M, Jault P, Ausset S, Auroy Y, et al. A serious game for massive training and assessment of French soldiers involved in forward combat casualty care (3D-SC1): development and deployment. *JMIR Serious Games*. 2016;4(1):e5.
50. Dankbaar ME, Alsmas J, Jansen EE, van Merriënboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract*. 2016;21(3):505–21.
51. Kaczmarczyk J, Davidson R, Bryden D, Haselden S, Vivekananda-Schmidt P. Learning decision making through serious games. *Clin Teach*. 2016;13(4):277–82.
52. Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med Teach*. 2013;35(1):e867–98.
53. Drummond D, Hadchouel A, Tesniere A. Serious games for health: three steps forwards. *Adv Simul (Lond)*. 2017;2:3.
54. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci*. 2014;111:8410–5.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



4.3. Conclusion

Cette 2nde étude a évalué l'efficacité de la simulation numérique avec *LabForGames Warning* sur l'apprentissage du raisonnement clinique en comparant un groupe d'étudiants infirmiers formé par simulation numérique par rapport à un groupe formé par un enseignement traditionnel. Le raisonnement clinique a été évalué par les tests de concordance de script immédiatement et 1 mois après une formation à la détection de l'aggravation clinique d'un patient. Cette étude multicentrique a inclus 146 étudiants infirmiers volontaires. Aucune différence significative n'a été observée sur le raisonnement clinique entre la formation par la simulation avec jeux sérieux et l'enseignement traditionnel. Cependant, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation numérique. La simulation avec jeux sérieux pourrait donc avoir un impact positif sur la formation à long terme au raisonnement clinique des infirmiers, mais cet impact reste à être exploré.

III- Discussion Générale

1. Motivation et jeux sérieux

Un élément important dans l'éducation est le temps passé au processus d'apprentissage : plus l'apprenant y passe du temps et répète l'exercice, meilleurs sont l'apprentissage et la rétention¹²⁵. Le jeu sérieux pourrait être un outil parfait quand on voit le temps que passent les joueurs de jeux vidéos devant leur écran. Mais pour cela, la motivation a un rôle important¹²⁶. Elle peut être définie comme le processus psychologique responsable du déclenchement, du maintien, de l'entretien ou de l'arrêt d'une conduite. Elle est l'ensemble des forces et des facteurs qui déterminent l'action et le comportement d'un individu pour atteindre un objectif ou réaliser une activité.

Selon Deci et Ryan, on peut distinguer plusieurs types de motivation chez les apprenants⁵⁹ : l'apprenant amotivé, l'apprenant motivé extrinsèquement, et celui motivé intrinsèquement. L'apprenant amotivé ne perçoit pas la valeur de l'activité d'apprentissage et ne croit pas que cela puisse donner une amélioration du résultat. Devant un jeu sérieux, le seul aspect qu'il aime est le côté environnement et son. Une fois cet effet fugace passé, il abandonne rapidement le jeu. L'apprenant extrinsèquement motivé n'aime pas l'activité d'apprentissage mais comprend que ça puisse améliorer son résultat. Un facteur externe va donc le motiver, comme obtenir une bonne note à l'examen après la formation. Sa difficulté est uniquement de trouver le courage d'apprendre. Enfin, l'apprenant motivé intrinsèquement aime l'activité d'apprentissage et en perçoit sa valeur. Dans son cas, le jeu sérieux est un outil parmi d'autres pour apprendre. Selon Drummond et al., pour les apprenants amotivés ou motivés intrinsèquement, le jeu sérieux en tant que tel a probablement moins d'intérêt par rapport à des méthodes d'enseignement conventionnelles⁵⁹. Par contre, le jeu sérieux pourrait avoir un intérêt important pour l'apprenant extrinsèquement motivé en l'aidant à trouver le courage d'apprendre et en lui montrant immédiatement un résultat tangible de cet apprentissage avec le score. Ces éléments de motivation intrinsèque et extrinsèque sont donc à prendre en compte dès le développement du jeu.

L'aspect « ludification » (« *gamification* ») du jeu sérieux apporte des facteurs motivationnels et des éléments cognitifs pour promouvoir l'apprentissage et garder des apprenants motivés et engagés^{126,127}. C'est d'ailleurs ce qui va différencier un jeu sérieux d'une simulation virtuelle sans ludification. Ces éléments de ludification peuvent être le score, des barres de progression, le challenge, l'imaginaire (Fantasy), des règles, des stimuli sensoriels, des connections sociales... Le challenge a un effet positif sur l'engagement de l'apprenant mais aussi sur l'apprentissage¹²⁸. Toutefois, il doit être adapté aux compétences de l'apprenant et doit pouvoir évoluer en fonction de son apprentissage, comme par exemple avec différents niveaux de difficulté possibles dans le jeu. Cependant, dans les études sur les jeux sérieux en santé, la description est parfois imprécise sur ces éléments. Dans notre jeu sérieux, les éléments

de ludification étaient le score et les appels téléphoniques différents au médecin en fonction des étapes précédemment effectuées par le joueur.

Des travaux antérieurs ont montré que les étudiants en santé sont très motivés à utiliser les jeux sérieux car ils les perçoivent comme plus engageants et plus interactifs que les méthodes traditionnelles et comme pouvant les aider dans l'apprentissage et la consolidation des connaissances^{72,92,129,130}. Cependant, l'équipe de Dankbaar et al. a observé que, bien que les étudiants soient plus motivés par le jeu sérieux par rapport à un e-module, l'effet en termes d'apprentissage n'a pas été retrouvé comme étant supérieur^{72,92}. Ces résultats sont semblables à ceux de notre étude mais il faut noter que l'apprentissage était évalué à court terme dans ces 2 études. Par contre, hors santé, une méta-analyse plus ancienne sur l'effet motivationnel des jeux sérieux ne retrouvait pas plus d'effet motivant dans le jeu sérieux qu'avec la méthode conventionnelle alors qu'ils retrouvaient un gain supérieur en termes d'apprentissage et de rétention¹³¹. Ces résultats pourraient s'expliquer par le fait que les mécanismes de jeu ne sont parfois qu'incomplètement intégrés dans les jeux sérieux, ou peut-être parce que l'usage du jeu sérieux par des apprenants non motivés ou déjà motivés intrinsèquement atténueraient la différence. Ainsi, dans des travaux de recherche futurs, il serait intéressant d'utiliser des échelles psychométriques évaluant plus précisément la motivation de l'apprenant. Par exemple, l'échelle de motivation situationnelle (*the Situational Motivation Scale, SIMS*) créée et validée par Guay et al.¹³² a pour objectif de déterminer quelle est la motivation et quel type de motivation régule les comportements d'un individu durant une activité. Ainsi, le profil de motivation des étudiants pourrait être évalué lors des sessions de formation avec le jeu et lors d'un enseignement traditionnel pour personnaliser la formation.

2. Jeux sérieux, raisonnement clinique, et stratégies pédagogiques

Le raisonnement clinique est un processus cognitif complexe. Comme tout processus cognitif, il nécessite un temps long d'apprentissage et de transfert. Le raisonnement clinique s'apprend sur la durée, sur des mois, et non pas en 1h de formation. Il est ainsi probablement illusoire de penser qu'une seule formation courte suffise pour améliorer l'apprentissage du raisonnement clinique chez des étudiants. Cet aspect ne favorise pas non plus l'évaluation du raisonnement clinique alors que les méthodes de son évaluation sont déjà peu standardisées. Par ailleurs, l'apprentissage du raisonnement se fait en général par des cas cliniques, mais l'apprentissage obtenu par un cas donné n'est pas nécessairement transférable aux autres cas ou à d'autres contextes¹³³. Norman et al. ont montré que l'apprentissage avec plusieurs exemples avait un effet important sur la capacité des étudiants à appliquer des concepts pour résoudre un nouveau problème¹³³. Middeke et al. se sont ainsi intéressés à la transférabilité du raisonnement clinique sur d'autres cas de jeu sérieux¹⁴. Des étudiants en médecine ont été exposés à un tronc commun de cas (>46 cas) par le jeu sérieux EMERGE sur la prise en charge de patients aux urgences. Sur 6 semaines, en plus du tronc commun, 3 groupes étaient randomisés : le groupe A avec 2 cas différents supplémentaires, le

groupe B avec 2 cas supplémentaires et différents de ceux du 1^{er} groupe et le groupe AB comportant les 4 cas précédents. A la fin, tous les étudiants passaient sur ces 4 cas du jeu pour être évalués au niveau de leur raisonnement clinique grâce à un score de jeu comportant différents items (histoire, examen clinique, examens complémentaires, traitement mis en place, diagnostic, transfert du patient...). Il n'a pas été retrouvé de différence significative entre les groupes. Les auteurs ont conclu que la formation au raisonnement clinique via le jeu sérieux n'était pas spécifique d'un cas mais pouvait être transférable à d'autres cas. Cependant une des limites discutées était l'absence de validation préalable de la construction de leur score via le jeu sérieux. Il n'est donc pas certain que leur score reflète le niveau de performance. En effet, notre étude a bien montré la difficulté de valider un score de jeux sérieux en terme de construit. Il n'y avait d'ailleurs pas de point négatif possible dans leur jeu, ce qui pouvait favoriser la minutie des étudiants sans différencier le raisonnement.

Un autre moyen probablement intéressant pour l'apprentissage du raisonnement clinique est la formation mixte (*blended learning*) c'est-à-dire associant plusieurs méthodes pédagogiques (présentielle et à distance)^{97,129}. Le jeu sérieux ne doit donc pas être un outil unique mais doit faire partie des outils s'intégrant dans une formation plus globale.

Certaines équipes se sont également intéressées à la possibilité de moduler la charge cognitive grâce au jeu sérieux^{84,134}. Dans l'étude de Mohan et al⁸⁴, 368 médecins urgentistes ont été randomisés en deux groupes : jeu sérieux vs programme éducationnel traditionnel pour la prise de décision au tri de patients polytraumatisés. Puis, une évaluation de leur performance au tri a été réalisée immédiatement après la formation et à 6 mois par simulation virtuelle, avec une modulation de la charge cognitive. Par rapport à l'éducation traditionnelle, l'exposition des médecins à un jeu sérieux a amélioré la prise de décision de tri dans une simulation virtuelle. Lorsqu'une charge cognitive élevée (diminution du nombre de zones d'accueil disponibles) était appliquée, aucune différence significative de décision de tri n'a été observée entre les groupes. Il n'est cependant pas exclu que la charge cognitive ne fonctionnait pas suffisamment bien, ou qu'elle ait entraîné plus de transfert et donc un meilleur tri dans le groupe traditionnel.

Une autre étude précédemment décrite avait comparé l'apprentissage du raisonnement clinique chez les étudiants en médecine selon 2 modalités pédagogiques différentes de jeu sérieux pour la prise en charge d'urgences abdominales⁸⁸. Dans une première modalité, les cas étaient nombreux et courts (éléments clés) et dans une seconde modalité, les cas étaient peu nombreux et longs (analyse systématique)⁸⁸. Aucune différence n'était retrouvée entre ces deux modalités.

Enfin, la revue de la littérature de Gorbanev et al. sur l'efficacité des jeux sérieux dans l'éducation médicale entre 2011 et 2015 (mais excluant le milieu paramédical et la pharmacie) a étudié les stratégies pédagogiques utilisées⁶⁸. Les 20 études retenues entraient dans les champs suivants : comportementaliste (n = 10 [47,6%]), cognitiviste (n = 6 [28,6%]), constructiviste (n = 3 [14,3%]) ou humaniste (n = 1 [4,8%]). Dans notre étude, la simulation par le jeu en associant le jeu sérieux à un débriefing avec un

enseignant semble utiliser une stratégie pédagogique plutôt humaniste. La phase d'expérience correspond au scénario puis la phase d'observation réflexive au débriefing, l'apprentissage étant centré sur l'apprenant, et l'instructeur étant utile comme facilitateur. Par contre, si l'on n'y associe pas cette phase de débriefing avec l'enseignant, notre jeu serait plutôt de nature comportementaliste, selon la théorie de l'instruction programmée. Dans cette dernière, le jeu est un modèle dans lequel l'apprenant progresse en jouant à sa propre vitesse, l'apprentissage étant favorisé par le feedback automatisé.

Ainsi le jeu sérieux peut, quand on travaille conjointement avec les équipes de développement, ouvrir des possibilités intéressantes de modulation pédagogique et de personnalisation de l'éducation dans le cadre de l'apprentissage du raisonnement clinique. Les travaux de recherche pédagogique sont à poursuivre dans ce domaine.

3. Raisonnement clinique et intelligence artificielle

Depuis plusieurs années, le développement de l'intelligence artificielle se poursuit dans le domaine médical^{51,52}. L'intelligence artificielle fait référence à des systèmes informatiques qui miment les fonctions cognitives humaines telles que l'apprentissage et la résolution de problème⁵¹. Ces systèmes informatiques ont la particularité d'apprendre à partir de données brutes avec un degré d'autonomie, comme cela se produit avec l'apprentissage automatique (*machine learning*: système réalisant une tâche sans instruction explicite), l'apprentissage en profondeur (*deep learning*: sous-catégorie de l'apprentissage automatique dans lequel le système informatique apprend et modélise des données en ajustant les variables selon un réseau articulé de neurones artificiels) et l'apprentissage par renforcement (*reinforcement learning*: consiste, pour un agent autonome (robot, etc.), à apprendre des actions, à partir d'expériences, de façon à optimiser au cours du temps la récompense). Dans le domaine de la santé, l'informatisation des données du patient permet de recueillir un volume important de données, ce qui est idéal pour des modèles d'intelligence artificielle et permet de donner des prédictions, des recommandations ou des diagnostics. Loftus et al. ont rappelé l'apport de l'intelligence artificielle dans les scores de prédiction chirurgicaux et son intérêt dans un futur proche dans les prises de décision chirurgicales⁵¹. L'intelligence artificielle est ainsi en train de « remplacer » l'être humain dans sa démarche diagnostique comme par exemple en radiologie, en dermatologie, mais aussi dans des situations cliniques plus complexes⁵². Ainsi, Liang et al ont développé un modèle utilisant l'intelligence artificielle pour le diagnostic de cas complexes de fièvre en pédiatrie. La preuve de concept a été confirmée par la haute précision des diagnostics qui étaient comparables à ceux des experts¹³⁵. L'interaction entre l'humain et la machine va donc être indispensable dans les années à venir dans toutes les démarches de raisonnement clinique.

Ces développements technologiques vont également probablement faire évoluer la compréhension de l'apprentissage du raisonnement clinique et de son évaluation. Par exemple, Hege et al.⁵⁰ ont développé un outil traduisant des éléments d'apprentissage

conceptuel du raisonnement clinique sur un logiciel. Les apprenants construisaient le script pathologique de leur patient, comme une carte conceptuelle, lorsqu'ils travaillaient sur un scénario de patient virtuel. Les auteurs ont alors observé que les cartes conceptuelles réalisées par les étudiants en médecine comportaient moins de connections que les experts, sans pouvoir pour autant en analyser les raisons. Dans le futur, ce type de logiciel explorant l'apprentissage conceptuel du raisonnement pourra probablement aider à mieux le comprendre. Toutes ces innovations ouvrent de nouvelles pistes de recherche technologique et/ou pédagogique.

4. Apport du débriefing dans le jeu sérieux

Le rôle du débriefing est fondamental dans l'apprentissage en simulation¹³⁶. Cette analyse réflexive du cas permet de comprendre la démarche de raisonnement en explorant l'ensemble des phases intuitives et analytiques. Le jeu sérieux n'étant qu'un outil pédagogique de simulation, le débriefing semble donc indissociable du jeu. Habituellement, des feedbacks automatisés pendant ou à la fin du jeu sont disponibles^{29,86,87}. Ces feedbacks de fin automatisés qui récapitulent le cas peuvent être assimilés à un débriefing automatisé. Cependant, l'intérêt d'un débriefing par des formateurs après une séance de jeu sérieux, sur le même modèle que celui de la simulation haute-fidélité, semble intéressant mais a été peu exploré⁶². Ce débriefing avec un instructeur permet de renforcer les objectifs pédagogiques du jeu et leurs apprentissages. Dans leur méta-analyse, Foronda et al. , soulignent que la présence du débriefing dans la simulation virtuelle est très variable⁷⁴. La plupart des études ne mentionnent que très rarement leurs pratiques bien que les étudiants soient souvent « débriefés par le jeu ». Comme dans notre étude, seules quelques équipes ont associé leur jeu à un débriefing avec un instructeur^{78,79,124}. Par exemple, Barré et al. ainsi que Michelet et al. ont mis en place, après chaque session de jeu, un débriefing avec un instructeur en ciblant des objectifs pédagogiques précis^{78,79}.

Cependant, la meilleure méthode de débriefing lors d'une simulation virtuelle ou d'un jeu sérieux n'est pas clairement établie. Dans l'étude de Chung et al., les participants ont évalué plus positivement la qualité du feedback dans le groupe patient standardisé ayant reçu un feedback avec un instructeur par rapport à celui automatisé délivré par le jeu sérieux⁹⁷. Cependant, l'objectif de l'étude n'était pas de comparer les deux modalités de débriefing. A contrario, Verkuyl et al. ont comparé 3 méthodes de débriefing après simulation virtuelle sur un jeu dans le domaine de la psychiatrie¹³⁷. Les étudiants infirmiers étaient randomisés en 3 groupes selon la méthode : auto-débriefing, débriefing avec instructeur et débriefing virtuel (via une plateforme web). Aucune différence significative n'a été retrouvée en termes de connaissance et d'auto-évaluation d'apprentissage entre les groupes. L'ensemble des étudiants ont donné de bons scores pour l'ensemble des débriefings avec cependant un moins bon score pour l'auto-débriefing. Dans cette étude, le débriefing virtuel se faisait via une plateforme web sur laquelle les étudiants et un instructeur se connectaient pour faire le débriefing. Mais qu'en est-il du débriefing présentiel avec un instructeur versus un débriefing

automatique via le jeu ? La présence ou non d'un débriefing dépend aussi des modalités d'utilisation du jeu sérieux : à visée formative ou évaluative, à distance ou non... Ainsi de nombreuses questions restent encore à explorer pour préciser la meilleure modalité du débriefing dans les jeux sérieux.

5. Le jeu sérieux du futur

Certaines limites technologiques restreignent encore des fonctionnalités et le réalisme. L'interrogatoire avec le patient se réduit souvent à un menu déroulant avec des questions prédéfinies comme dans notre propre jeu. Le développement de l'analyse conversationnelle est en plein essor dans le domaine de la santé mais reste encore sous forme d'une preuve de concept¹³⁸. Dans le cadre d'un projet ANR intitulé « LabForSIMS 2 », notre équipe travaille avec deux équipes du CEA pour intégrer l'analyse conversationnelle dans un jeu portant sur une urgence chirurgicale et un autre sur la réanimation du nouveau-né¹³⁹. Des tests ont été réalisés pour évaluer la fiabilité du système^{139,140}. De plus, le développement de la réalité mixte qui associe un mannequin tangible à la réalité virtuelle permettrait d'augmenter le réalisme et les possibilités pédagogiques. En effet, le mannequin tangible a certaines limites comme l'impossibilité de changer de couleur, d'avoir des mouvements fins ; d'un autre côté, la réalité virtuelle ne permet pas de toucher, ni de faire des gestes avec du vrai matériel. Notre projet LabForSIMS 2 travaille également sur ce développement technologique¹⁴¹. Ainsi, le développement technologique repousse les limites des outils pédagogiques à disposition des enseignants.

Conclusion

Cette thèse a permis une analyse de la place du jeu sérieux dans la formation au raisonnement clinique des professionnels de santé. Nous avons tout d'abord confirmé la validité du jeu sérieux *LabForGames Warning* en tant qu'outil pédagogique à visée formative et non sommative. Puis, bien qu'aucune différence d'apprentissage du raisonnement clinique n'ait été observée entre la formation par la simulation avec jeu sérieux et l'enseignement traditionnel, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation numérique. D'autres études sont à poursuivre pour préciser les modalités et la valeur éducative des jeux sérieux dans la formation des professionnels de santé, comme par exemple la place du débriefing.

Références

References

1. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf.* 2017;26(6):484-494.
2. Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf.* 2014;23(9):727-731.
3. Stiegler MP, Gaba DM. Decision-making and cognitive strategies. *Simul Healthc.* 2015;10(3):133-138.
4. Cresswell KM, Panesar SS, Salvilla SA, et al. Global research priorities to better understand the burden of iatrogenic harm in primary care: an international Delphi exercise. *PLoS Med.* 2013;10(11):e1001554.
5. Custer JW, Winters BD, Goode V, et al. Diagnostic errors in the pediatric and neonatal ICU: a systematic review. *Pediatr Crit Care Med.* 2015;16(1):29-36.
6. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med.* 2011;86(3):307-313.
7. Abimanyi-Ochom J, Bohingamu Mudiyansele S, Catchpool M, Firipis M, Wanni Arachchige Dona S, Watts JJ. Strategies to reduce diagnostic errors: a systematic review. *BMC Med Inform Decis Mak.* 2019;19(1):174.
8. Thammasitboon S, Cutrer WB. Diagnostic decision-making and strategies to improve diagnosis. *Curr Probl Pediatr Adolesc Health Care.* 2013;43(9):232-241.
9. Audetat MC, Laurin S, Dory V, Charlin B, Nendaz MR. Diagnosis and management of clinical reasoning difficulties: Part I. Clinical reasoning supervision and educational diagnosis. *Med Teach.* 2017;39(8):792-796.
10. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ.* 2010;44(1):94-100.
11. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Acad Med.* 2017;92(1):23-30.
12. Audetat MC, Laurin S, Dory V, Charlin B, Nendaz MR. Diagnosis and management of clinical reasoning difficulties: Part II. Clinical reasoning difficulties: Management and remediation strategies(). *Med Teach.* 2017;39(8):797-801.
13. Lambe KA, O'Reilly G, Kelly BD, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf.* 2016;25(10):808-820.
14. Middeke A, Anders S, Schuelper M, Raupach T, Schuelper N. Training of clinical reasoning with a Serious Game versus small-group problem-based learning: A prospective study. *PLoS One.* 2018;13(9):e0203851.
15. Middeke A, Anders S, Raupach T, Schuelper N. Transfer of Clinical Reasoning Trained With a Serious Game to Comparable Clinical Problems: A Prospective Randomized Study. *Simul Healthc.* 2020.
16. Chon SH, Timmermann F, Dratsch T, et al. Serious Games in Surgical Medical Education: A Virtual Emergency Department as a Tool for Teaching Clinical Reasoning to Medical Students. *JMIR Serious Games.* 2019;7(1):e13028.
17. Pelaccia T, Tardif J, Triby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online.* 2011;16.

18. Simmons B. Clinical reasoning: concept analysis. *J Adv Nurs*. 2010;66(5):1151-1158.
19. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022-1028.
20. Levett-Jones T, Hoffman K, Dempsey J, et al. The 'five rights' of clinical reasoning: an educational model to enhance nursing students' ability to identify and manage clinically 'at risk' patients. *Nurse Educ Today*. 2010;30(6):515-520.
21. Lee J, Lee YJ, Bae J, Seo M. Registered nurses' clinical reasoning skills and reasoning process: A think-aloud study. *Nurse Educ Today*. 2016;46:75-80.
22. Flin R, Maran N. Basic concepts for crew resource management and non-technical skills. *Best Pract Res Clin Anaesthesiol*. 2015;29(1):27-39.
23. Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth*. 2002;88(3):418-429.
24. Stiegler MP, Tung A. Cognitive processes in anesthesiology decision making. *Anesthesiology*. 2014;120(1):204-217.
25. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf*. 2013;22 Suppl 2:ii58-ii64.
26. Yazdani S, Hoseini Abardeh M. Five decades of research and theorization on clinical reasoning: a critical review. *Adv Med Educ Pract*. 2019;10:703-716.
27. Pelaccia T, Tardif J, Tribby E, et al. Insights into emergency physicians' minds in the seconds before and into a patient encounter. *Intern Emerg Med*. 2015;10(7):865-873.
28. Klein G. Naturalistic decision making. *Hum Factors*. 2008;50(3):456-460.
29. Kleinert R, Heiermann N, Plum PS, et al. Web-Based Immersive Virtual Patient Simulators: Positive Effect on Clinical Reasoning in Medical Education. *J Med Internet Res*. 2015;17(11):e263.
30. Simmons B, Lanuza D, Fonteyn M, Hicks F, Holm K. Clinical reasoning in experienced nurses. *West J Nurs Res*. 2003;25(6):701-719; discussion 720-704.
31. Hogan MP, Pace DE, Hapgood J, Boone DC. Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. *J Trauma*. 2006;61(5):1047-1052.
32. Charlin B, Bordage G, Van Der Vleuten C. L' évaluation du raisonnement clinique. *Pédagogie Médicale*. 2003;4(1):42-52.
33. Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association between licensing examination scores and resource use and quality of care in primary care practice. *JAMA*. 1998;280(11):989-996.
34. Liaw SY, Scherpbier A, Klainin-Yobas P, Rethans JJ. Rescuing A Patient In Deteriorating Situations (RAPIDS): an evaluation tool for assessing simulation performance on clinical deterioration. *Resuscitation*. 2011;82(11):1434-1439.
35. Nayer M, Glover Takahashi S, Hrynchak P. Twelve tips for developing key-feature questions (KFQ) for effective assessment of clinical reasoning. *Med Teach*. 2018;40(11):1116-1122.
36. Hrynchak P, Takahashi SG, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ*. 2014;48(9):870-883.
37. Karila L, Francois H, Monnet X, et al. [The Script Concordance Test: A multimodal teaching tool]. *Rev Med Interne*. 2018;39(7):566-573.
38. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak*. 2008;8:18.

39. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CP. Script concordance testing: a review of published validity evidence. *Med Educ*. 2011;45(4):329-338.
40. Sibert L, Charlin B, Corcos J, Gagnon R, Grise P, van der Vleuten C. Stability of clinical reasoning assessment results with the Script Concordance test across two different linguistic, cultural and learning environments. *Med Teach*. 2002;24(5):522-527.
41. Deschenes MF, Charlin B, Gagnon R, Goudreau J. Use of a script concordance test to assess development of clinical reasoning in nursing students. *J Nurs Educ*. 2011;50(7):381-387.
42. Deschênes MF, Goudreau J. Addressing the development of both knowledge and clinical reasoning in nursing through the perspective of script concordance: an integrative literature review. *J Nurs Educ*. 2017;7(12):28-38.
43. Koivisto JM, Multisilta J, Niemi H, Katajisto J, Eriksson E. Learning by playing: A cross-sectional descriptive study of nursing students' experiences of learning clinical reasoning. *Nurse Educ Today*. 2016;45:22-28.
44. Liou SR, Liu HC, Tsai HM, et al. The development and psychometric testing of a theory-based instrument to evaluate nurses' perception of clinical reasoning competence. *J Adv Nurs*. 2016;72(3):707-717.
45. Liaw SY, Rashasegaran A, Wong LF, et al. Development and psychometric testing of a Clinical Reasoning Evaluation Simulation Tool (CREST) for assessing nursing students' abilities to recognize and respond to clinical deterioration. *Nurse Educ Today*. 2018;62:74-79.
46. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1:31.
47. Fleiszer D, Hoover ML, Posel N, Razek T, Bergman S. Development and Validation of a Tool to Evaluate the Evolution of Clinical Reasoning in Trauma Using Virtual Patients. *J Surg Educ*. 2018;75(3):779-786.
48. Cook RJ, Durning SJ. Clinical Process Modeling: An Approach for Enhancing the Assessment of Physicians' Clinical Reasoning. *Acad Med*. 2019;94(9):1317-1322.
49. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ*. 2012;46(5):454-463.
50. Hege I, Kononowicz AA, Adler M. A Clinical Reasoning Tool for Virtual Patients: Design-Based Research Study. *JMIR Med Educ*. 2017;3(2):e21.
51. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial Intelligence and Surgical Decision-Making. *JAMA Surg*. 2019.
52. Pinnock R, McDonald J, Ritchie D, Durning SJ. Humans and machines: Moving towards a more symbiotic approach to learning clinical reasoning. *Med Teach*. 2019:1-6.
53. Granry JC. Guide de bonnes pratiques en matière de simulation en santé. 2012; https://www.has-sante.fr/upload/docs/application/pdf/2013-01/guide_bonnes_pratiques_simulation_sante_guide.pdf.
54. 2020 Healthcare Simulation Dictionary—Second Edition. 2020; <https://www.ahrq.gov/patient-safety/resources/simulation/terms.html>.
55. Cant R, Cooper S, Sussex R, Bogossian F. What's in a Name? Clarifying the Nomenclature of Virtual Simulation. *Clinical Simulation in Nursing*. 2019;27:26-30.
56. Wang R, DeMaria S, Jr., Goldberg A, Katz D. A Systematic Review of Serious Games in Training Health Care Professionals. *Simul Healthc*. 2016;11(1):41-51.

57. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. *J Med Internet Res*. 2019;21(3):e12994.
58. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg*. 2012;99(10):1322-1330.
59. Drummond D, Hadchouel A, Tesniere A. Serious games for health: three steps forwards. *Adv Simul (Lond)*. 2017;2:3.
60. Freeman S, Eddy SL, McDonough M, et al. Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci*. 2014;111:8410-8415.
61. Sitzmann T. A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Pers Psychol*. 2011;64:89-528.
62. Crookall D. Serious games, debriefing, and simulation/gaming as a discipline *Simulation & Gaming*. 2010;41(6):898-920.
63. Kolb AY, Kolb DA. Learning styles and learning spaces: enhancing experiential learning in higher education. *Acad Manag Learn Educ*. 2005;4:193-212.
64. Weidman J, Baker K. The Cognitive Science of Learning: Concepts and Strategies for the Educator and Learner. *Anesth Analg*. 2015;121(6):1586-1599.
65. Rouse DN. Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspect Health Inf Manag*. 2011;8:1c.
66. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Acad Med*. 2010;85(10):1589-1602.
67. Kononowicz AA, Woodham LA, Edelbring S, et al. Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J Med Internet Res*. 2019;21(7):e14676.
68. Gorbanev I, Agudelo-Londono S, Gonzalez RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online*. 2018;23(1):1438718.
69. Sipiyyaruk K, Gallagher JE, Hatzipanagos S, Reynolds PA. A rapid review of serious games: From healthcare education to dental education. *Eur J Dent Educ*. 2018;22(4):243-257.
70. Boeker M, Andel P, Vach W, Frankenschmidt A. Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. *PLoS One*. 2013;8(12):e82328.
71. Drummond D, Delval P, Abdenouri S, et al. Serious game versus online course for pretraining medical students before a simulation-based mastery learning course on cardiopulmonary resuscitation: A randomised controlled study. *Eur J Anaesthesiol*. 2017;34(12):836-844.
72. Dankbaar ME, Richters O, Kalkman CJ, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. *BMC Med Educ*. 2017;17(1):30.
73. De Sena DP, Fabricio DD, da Silva VD, Bodanese LC, Franco AR. Comparative evaluation of video-based on-line course versus serious game for training medical students in cardiopulmonary resuscitation: A randomised trial. *PLoS One*. 2019;14(4):e0214722.

74. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual Simulation in Nursing Education: A Systematic Review Spanning 1996 to 2018. *Simul Healthc*. 2020;15(1):46-54.
75. Kyaw BM, Saxena N, Posadzki P, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J Med Internet Res*. 2019;21(1):e12959.
76. Bracq MS, Michinov E, Jannin P. Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals: A Systematic Review. *Simul Healthc*. 2019;14(3):188-194.
77. Youngblood P, Harter PM, Srivastava S, Moffett S, Heinrichs WL, Dev P. Design, development, and evaluation of an online virtual emergency department for training trauma teams. *Simul Healthc*. 2008;3(3):146-153.
78. Barré J, Michelet D, Job A, et al. Does repeated exposure to critical situations in a screen-based simulation improve the self-assessment of non-technical skills in postpartum hemorrhage management? . *Simul Gaming*. 2019.
79. Michelet D, Barre J, Job A, et al. Benefits of Screen-Based Postpartum Hemorrhage Simulation on Nontechnical Skills Training: A Randomized Simulation Study. *Simul Healthc*. 2019;14(6):391-397.
80. Posel N, McGee JB, Fleiszer DM. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Med Teach*. 2015;37(9):813-818.
81. Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebo A. A Serious Game for Teaching Nursing Students Clinical Reasoning and Decision-Making Skills. *Stud Health Technol Inform*. 2016;225:905-906.
82. Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebo A. Nursing students' perceptions of a video-based serious game's educational value: A pilot study. *Nurse Educ Today*. 2018;62:62-68.
83. Kleinert R, Heiermann N, Wahba R, Chang DH, Holscher AH, Stippel DL. Design, Realization, and First Validation of an Immersive Web-Based Virtual Patient Simulator for Training Clinical Decisions in Surgery. *J Surg Educ*. 2015;72(6):1131-1138.
84. Mohan D, Farris C, Fischhoff B, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ*. 2017;359:j5416.
85. Knight JF, Carley S, Tregunna B, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. *Resuscitation*. 2010;81(9):1175-1179.
86. Bogossian FE, Cooper SJ, Cant R, Porter J, Forbes H, Team FAR. A trial of e-simulation of sudden patient deterioration (FIRST2ACT WEB) on student learning. *Nurse Educ Today*. 2015;35(10):e36-42.
87. Liaw SY, Chan SW, Chen FG, Hooi SC, Siau C. Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: randomized controlled trial. *J Med Internet Res*. 2014;16(9):e214.
88. Schubach F, Goos M, Fabry G, Vach W, Boeker M. Virtual patients in the acquisition of clinical reasoning skills: does presentation mode matter? A quasi-randomized controlled trial. *BMC Med Educ*. 2017;17(1):165.
89. LeFlore JL, Anderson M, Zielke MA, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. *Simul Healthc*. 2012;7(1):10-17.

90. McGrath J, Kman N, Danforth D, et al. Virtual alternative to the oral examination for emergency medicine residents. *West J Emerg Med.* 2015;16(2):336-343.
91. Lin DT, Park J, Liebert CA, Lau JN. Validity evidence for Surgical Improvement of Clinical Knowledge Ops: a novel gaming platform to assess surgical decision making. *Am J Surg.* 2015;209(1):79-85.
92. Dankbaar ME, Alsma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract.* 2016;21(3):505-521.
93. Liaw SY, Chng DYJ, Wong LF, et al. The impact of a Web-based educational program on the recognition and management of deteriorating patients. *J Clin Nurs.* 2017;26(23-24):4848-4856.
94. Graafland M, Bemelman WA, Schijven MP. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. *Surg Endosc.* 2017;31(10):4093-4101.
95. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an Electronic Game for Medical Education on Insulin Therapy: A Randomized Controlled Trial With Primary Care Physicians. *J Med Internet Res.* 2017;19(3):e72.
96. Mohan D, Fischhoff B, Angus DC, et al. Serious games may improve physician heuristics in trauma triage. *Proc Natl Acad Sci U S A.* 2018;115(37):9204-9209.
97. Chung C, Cooper SJ, Cant RP, et al. The educational impact of web-based and face-to-face patient deterioration simulation programs: An interventional trial. *Nurse Educ Today.* 2018;64:93-98.
98. Haerling KA. Cost-Utility Analysis of Virtual and Mannequin-Based Simulation. *Simul Healthc.* 2018;13(1):33-40.
99. Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. *Resuscitation.* 2004;62(2):137-141.
100. Gao H, McDonnell A, Harrison DA, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med.* 2007;33(4):667-679.
101. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg.* 2009;250(6):1029-1034.
102. Mackintosh N, Sandall J. Overcoming gendered and professional hierarchies in order to facilitate escalation of care in emergency situations: the role of standardised communication protocols. *Soc Sci Med.* 2010;71(9):1683-1686.
103. Nagpal K, Arora S, Vats A, et al. Failures in communication and information transfer across the surgical care pathway: interview study. *BMJ Qual Saf.* 2012;21(10):843-849.
104. SAED-un guide pour faciliter la communication entre professionnels de sante. http://www.has-sante.fr/portail/jcms/c_1776178/fr/saed-un-guide-pour-faciliter-la-communication-entre-professionnels-de-sante. Accessed 03-17- 2019.
105. Legifrance.gouv. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037218115&categorieLien=id>. Accessed 03-17-2020.
106. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med.* 2013;88(6):872-883.

107. Graafland M, Dankbaar M, Mert A, et al. How to systematically assess serious games applied to health care. *JMIR Serious Games*. 2014;2(2):e11.
108. Graafland M, Vollebergh MF, Lagarde SM, van Haperen M, Bemelman WA, Schijven MP. A serious game can be a valid method to train clinical decision-making in surgery. *World J Surg*. 2014;38(12):3056-3062.
109. Sugand K, Mawkin M, Gupte C. Training effect of using Touch Surgery for intramedullary femoral nailing. *Injury*. 2016;47(2):448-452.
110. Chaudery M, Clark J, Dafydd DA, et al. The Face, Content, and Construct Validity Assessment of a Focused Assessment in Sonography for Trauma Simulator. *J Surg Educ*. 2015;72(5):1032-1038.
111. Kowalewski KF, Hendrie JD, Schmidt MW, et al. Validation of the mobile serious game application Touch Surgery for cognitive training and assessment of laparoscopic cholecystectomy. *Surg Endosc*. 2017;31(10):4058-4066.
112. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract*. 2014;19(2):233-250.
113. Messick S. *Validity*. 3rd ed. New York, NY: American Council on education and Macmillan. Cité par Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1:31.; 1989.
114. Borgersen NJ, Naur TMH, Sorensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Ann Surg*. 2018;267(6):1063-1068.
115. Gerard JM, Scalzo AJ, Borgman MA, et al. Validity Evidence for a Serious Game to Assess Performance on Critical Pediatric Emergency Medicine Scenarios. *Simul Healthc*. 2018;13(3):168-180.
116. Kane MT. *Educational measurement*. 4th ed: Praeger; 2006.
117. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575.
118. Hillman KM, Bristow PJ, Chey T, et al. Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med*. 2002;28(11):1629-1634.
119. Chua WL, See MTA, Legio-Quigley H, Jones D, Tee A, Liaw SY. Factors influencing the activation of the rapid response system for clinically deteriorating patients by frontline ward clinicians: a systematic review. *Int J Qual Health Care*. 2017;29(8):981-998.
120. Liaw SY, Scherpbier A, Klainin-Yobas P, Rethans JJ. A review of educational strategies to improve nurses' roles in recognizing and responding to deteriorating patients. *Int Nurs Rev*. 2011;58(3):296-303.
121. Connell CJ, Endacott R, Jackman JA, Kiprillis NR, Sparkes LM, Cooper SJ. The effectiveness of education in the recognition and management of deteriorating patients: A systematic review. *Nurse Educ Today*. 2016;44:133-145.
122. Liaw SY, Wong LF, Lim EY, et al. Effectiveness of a Web-Based Simulation in Improving Nurses' Workplace Practice With Deteriorating Ward Patients: A Pre- and Postintervention Study. *J Med Internet Res*. 2016;18(2):e37.
123. Cooper S, Kinsman L, Buykx P, McConnell-Henry T, Endacott R, Scholes J. Managing the deteriorating patient in a simulated environment: nursing students' knowledge, skill and situation awareness. *J Clin Nurs*. 2010;19(15-16):2309-2318.
124. Blanie A, Amorim MA, Benhamou D. 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 Med Educ*. 2020;20(1):53.

125. Cook DA, Hamstra SJ, Brydges R, et al. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med Teach*. 2013;35(1):e867-898.
126. Morris BJ, Croker S, Zimmerman C, Gill D, Romig C. Gaming science: the "Gamification" of scientific thinking. *Front Psychol*. 2013;4:607.
127. Sardi L, Idri A, Fernandez-Aleman JL. A systematic review of gamification in e-Health. *J Biomed Inform*. 2017;71:31-48.
128. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. *Comput Hum Behav*. 2016;54:170-179.
129. Dankbaar M. Serious games and blended learning; effects on performance and motivation in medical education. *Perspect Med Educ*. 2016.
130. Kaczmarczyk J, Davidson R, Bryden D, Haselden S, Vivekananda-Schmidt P. Learning decision making through serious games. *Clin Teach*. 2016;13(4):277-282.
131. Wouters P, Van Nimwegen C, Van Oostendorp H, Van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*. 2013;105(2):249-265.
132. Guay F, Vallerand RJ, Blanchard C. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*. 2000;24(3):175-213.
133. Norman G, Dore K, Krebs J, Neville AJ. The power of the plural: effect of conceptual analogies on successful transfer. *Acad Med*. 2007;82(10 Suppl):S16-18.
134. Mohan D, Angus DC, Ricketts D, et al. Assessing the validity of using serious game technology to analyze physician decision making. *PLoS One*. 2014;9(8):e105445.
135. Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. 2019;25(3):433-438.
136. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc*. 2007;2(2):115-125.
137. Verkuyl M, Attack L, McCulloch T, et al. Comparison of Debriefing Methods after a Virtual Simulation: An Experiment. *Clinical Simulation in Nursing*. 2018;19:1-7.
138. Maicher K, Danforth D, Price A, et al. Developing a Conversational Virtual Standardized Patient to Enable Students to Practice History-Taking Skills. *Simul Healthc*. 2017;12(2):124-131.
139. Laleye F, De Chalendar G, Blanié A, Brouquet A, D. B. A French Medical Conversations Corpus Annotated for a Virtual Patient Dialogue System. LREC; 2019.
140. Laleye F, Blanié A, Brouquet A, D. B, De Chalendar G. Hybridation d'un agent conversationnel avec des plongements lexicaux pour la formation au diagnostic médical. *Traitement Automatique des Langues Naturelles (TALN)*; 2019.
141. Prokopenko K, Dupont R. Towards Dense 3D Reconstruction for Mixed Reality in Healthcare: Classical Multi-View Stereo vs Deep Learning. *International Conference on Computer Vision (ICCV 2019)*; 2019; Seoul.

Annexe 1 : Curriculum Vitae

Antonia BLANIE

(ex épouse Odri)

Née le 06-02-80 à Paris

53 rue de l'amiral Mouchez 75013 Paris

0683222580 - antonia.blanie@aphp.fr

Inscription à l'ordre des médecins 94/16624—RPPS : 10100171734

ORCID : 0000-0003-2291-0315

Activité- Formation Médicale

- Depuis le 1^{er} novembre 2010 : Anesthésiste Réanimateur - Département d'Anesthésie Réanimation (Pr Duranteau), au CHU de Bicêtre.
 - **Praticien hospitalier temps plein** (2015-)
 - **Chef de Clinique Assistant** (2011-2014)
 - **Praticien hospitalier contractuel** (2010-2011)
- Activité en anesthésie adulte et gardes en anesthésie adulte et réanimation chirurgicale
- Coordonnateur adjoint de l'unité de chirurgie ambulatoire (2019-)
- 2004-10 : TCEM du DES d'Anesthésie-Réanimation à Paris.
 - **Diplôme de DES d'Anesthésie Réanimation** (Paris 5, 2010). « Durée de la fibrinolyse postopératoire après prothèse totale de hanche et genou : étude préliminaire » (Pr Ozier et Dr Rosencher, CHU Cochin).
 - **Diplôme de Docteur en Médecine** (Paris 5, 2010) « Peut-on utiliser la succinylcholine en réanimation ? Etude prospective et observationnelle » (Dr Vigué, CHU de Bicêtre).
- 2003-04 : Examen National Classant (534^{ème}).
- 1997-2004 : PCEM et DCEM, à l'Université Paris 5, Faculté Cochin Port-Royal.
- 1996-97 : Baccalauréat scientifique, spécialité mathématiques, Meudon (92)

Autres diplômes universitaires et parcours scientifique

- **Membre du comité d'analyse et maîtrise du risque (CAMR)** de la SFAR (2017-)
- **Doctorat en cours à l'ED566 de l'université Paris Saclay, CIAMS EA 4532** (Co-directeur : Pr Amorim et Pr Benhamou) (2016-)
- **DU formateur à l'enseignement de la médecine sur simulateur** (Paris 5) (2013-14)
- **DIU d'échocardiographie d'île de France** (Pr Gueret, Paris 12) (2007-10)
- **Master 2 « Sciences, santé et applications », Mention « biologie cellulaire, physiologie et pathologie », Spécialité « cardiovasculaire »** Paris 7 cohabilitée avec Paris 5. Dr Vigué et Dr Geeraerts. « Hétérogénéité de la vulnérabilité cérébrale à

l'hypoxie-hypotension après traumatisme crânien chez le rat: comparaison cortex et noyaux gris centraux» (2008-09)

- **Master 1 Santé parcours Physiologie** (Paris 5) (2006-2007)
- **Certificats de MSBM:** Différenciation et Oncogenèse, (2000-01, Paris 6) ; Physiologie et Biologie des Systèmes Intégrés (Paris 5) (1999-2000)

Activités en simulation en santé et pédagogie

- **Responsable adjointe du centre de simulation LabForSIMS** de Faculté de Médecine Paris Sud (2016-) et responsable pédagogique-scientifique (2011-).
- **Formateur :** Qualification de formateur par :
 - 2013-14: **DU de formateur à l'enseignement de la médecine sur simulateur** (Paris 5). Mémoire sur « mise en place d'un jeu sérieux en médecine »
 - **Activité régulière de formateur** à LabForSIMS depuis 2011 : sur simulateur haute fidélité à LabForSIMS. (Par ex : module obligatoire simulation des DESAR d'Ile de France depuis 2011), simulateur basse fidélité et Jeux sérieux.
 - **Instructeur au workshop** de simulation haute-fidélité aux congrès de l'ESA (European Society of anesthesiology) (2017-).
 - **Instructeur au Bloc 360°** congrès SFAR 2019
 - **Responsable du workshop** de simulation haute-fidélité au congrès du MAPAR (2018-)
 - Formation « debriefing with good judgment » workshop 2018
- **Responsable et enseignant dans la « Formation aux gestes élémentaires** pour la prise de garde » obligatoire pour les DFASM1 de la Faculté de Médecine Paris Sud, UE 13 stage et garde (ateliers VVP, PL...)
- **Responsable pédagogique de la mise en place de jeux sérieux** (développement et intégration dans le cursus de formation)
 - Pour étudiants en médecine: pathologies neurologiques aiguës aux urgences (APP innovation pédagogique Université Paris Sud 2014)
 - Pour les étudiants en soins infirmiers: détection de l'aggravation d'un patient et la communication interprofessionnelle avec l'outil SAED (APP ARS simulation en santé 2015).
 - Pour les urgences chirurgicales dans le cadre d'un projet IDEFI-N 2014 de l'ANR
- **Responsable adjointe et enseignant à la formation de formateur courte en simulation en santé** organisée par LabForSIMS-Faculté de médecine Paris Sud (2015-).
- **Coordinatrice de l'Audit externe de LabForSIMS** par la SOFRASIMS-HAS en 2016
- **Membre du Projet LabForSIMS 2 : IDEFI-Numérique 2014 ANR** en collaboration avec CEA et HRV.
- **Au sein de la SoFraSimS (Société francophone de simulation en Santé) :**
 - **Membre du CA de la SoFraSimS** (2017-)
 - **Membre du groupe de recherche de la Sofrasims** (2017-)

- Membre du groupe de travail sur « formation courte de formateur en simulation »
- **Membre de la Commission d'innovation pédagogique** de la Faculté de médecine Paris Sud (2017-)
- **Enseignements au sein du DU « Formation en simulation et Immersion »** (Université de Lyon) **et différents congrès** (voir communications)
- Membre du groupe de travail sur les recommandations pour les pratiques professionnels SFAR/SRFLF/SOFRASIMS/SFMU « intérêt de l'apprentissage par la simulation en soins critiques » (2019)

Activités de Recherche

- **Co investigateur** : Etude CatCH (2011); Etude OCTOPUS (2014) ; Etude IRIS: (2014) ; Observatoire national sur les NACO et gestes invasifs/hémorragies aiguës (2015), enquête sur le sommeil avant séance de simulation haute fidélité (2017)
- **Investigateur principal de plusieurs études en pédagogie et simulation**
- **Encadrement de mémoires et thèse depuis 2015** (élèves IADE, IADE, internes DESAR)
 - « Fiabilité du Nexfin® et du doppler oesophagien en chirurgie majeure non cardiaque » Mémoire interne DESAR 2015 (40h)
 - « Rétention des connaissances sur la formation des gestes d'urgences » Mémoire recherche IADE en 2015 (10h)
 - « Evaluation des connaissances des internes d'anesthésie réanimation sur les compétences non techniques et gestion des risques. Mémoire recherche IADE en 2016 (10h)
 - « Labforgames : premières évaluations de l'outil pédagogique chez les étudiants en médecine. Mémoire recherche IADE en 2016 (10h)
 - « Comparaison de l'apprentissage des connaissances en fonction du rôle de l'apprenant (acteur vs spectateur) en simulation haute-fidélité ». Mémoire interne DESAR 2016 (40h)
 - « Effet du rôle de l'apprenant (acteur-spectateur ou spectateur) sur l'apprentissage lors de sessions de simulation haute-fidélité chez des internes d'anesthésie-réanimation : étude monocentrique, prospective et randomisée ». Mémoire interne DESAR 2017 (40h)
 - « Impact sur l'apprentissage des compétences techniques de l'utilisation d'une grille d'observation chez des internes d'anesthésie-réanimation observateurs en séance de simulation : étude randomisée ». Thèse interne DESAR (2017 (40h) (Directeur de thèse).
 - « Enquête sur la connaissance et l'utilisation des aides cognitives de crise en anesthésie réanimation en France en 2017 ». Thèse interne DESAR (2017 (40h) (Directeur de thèse)
 - « Impact sur l'apprentissage des compétences non techniques de l'utilisation d'une grille d'observation non technique chez des internes d'anesthésie-réanimation observateurs en séance de simulation par rapport à un groupe sans support. » Mémoire d'interne DESAR en 2019 (40h)
 - « Utilisation des aides cognitives dans 5 scénarios de crise en simulation en réanimation : analyse observationnelle à partir d'enregistrement vidéo. » Thèse interne DESAR (2019 (40h) (Directeur de thèse)
- **Reviewer** pour ACCPM dans le domaine de la simulation en 2016

- Membre du comité scientifique du groupe de recherche de la Sofrasims (2017-)
- **Projet LabForSIMS 2 : IDEFI-Numérique 2014 ANR** en collaboration avec CEA et HRV.

Publications scientifiques (auteurs ou investigateurs)

Score SIGAPS en 2020 = 339

- Blanié A, Penon C, Edouard A, Benhamou D. Asystolie lors d'une laryngoscopie, Ann Fr Anesth Reanim. 2006 Aug;25(8):904.
- Thévenin A, Beloeil H, Blanié A, Benhamou D, Mazoit JX. The limited efficacy of tramadol in postoperative patients: a study of ED80 using the continual reassessment method. Anesth Analg. 2008 Feb;106(2):622-7
- Odri A, Cavalcanti A, Billard V. Un indice bispectral étrangement élevé. Ann Fr Anesth Reanim. 2008 nov;27(11): 945–948
- Odri A, Geeraerts T, Vigué B. Hypothermie et protection cérébrale après traumatisme crânien- Influence des gaz du sang. Club de l'ANARLF. Ann Fr Anesth Reanim. 2009;28(4) :352–357.
- Blanié A, Fadel E, Duranteau J. Left pulmonary artery transection following penetrating thoracic trauma. Journal of Trauma. 2011 Nov;71(5):1479.
- Blanié A, Leblanc PE, Ract C, Pottecher J, Cheisson G, Huet O, Laplace C, Lopes T, Duranteau J, Vigué B. The limits of succinylcholine for critically ill Patients. Anesth Analg. 2012 Oct; 115(4): 873-9
- Blanié A, Vigué B, Benhamou D, Duranteau J, Geeraerts T. Frontal lobe and thalamus have different sensitivities to hypoxia-hypotension after traumatic brain injury: a microdialysis study in rats. J Neurotrauma. 2012 Dec; 29(18): 2782-90
- Blanié A, Bellamy L, Rhayem Y, Flaujac C, Samama CM, Fontenay M, Rosencher N. Duration of postoperative fibrinolysis after total hip or knee replacement: a laboratory follow up study. Thromb Res. 2013 Jan;131(1):e6-e11.
- Blanié A, Ract C, Mazoit JX, Vigué B. In response. Anesth Analg. 2013 Mai; 116(5): 1182-3.
- Blanié A, Rosencher N, Godier A. Antifibrinolytic thereapy. Thromb Res. 2014 Jan;133(1):127
- Blanié A, Soued M, Mazoit JXM, Benhamou D, Duranteau J. Comparison of photoplethysmography vs. esophageal Doppler for the assessment of cardiac index during major non-cardiac surgery. Anesth Analg. 2016 Feb;122(2):430-6
- Blanie A, Benhamou D. Can linear cognitive aids always be used in anaesthesia? Anaesthesia. 2016 Nov;71(11):1377-1378.
- Benhamou D, Blanié A. Méthodes de mesure du travail en équipe. Institut pour la qualité et la sécurité en santé. Séminaire Recherche – Paris – 23 juin 2016
- Albaladejo P, Samama CM, Sié P, Kauffmann S, Mémier V, Suchon P, Viallon A, David JS, Gruel Y, Bellamy L, de Maistre E, Romegoux P, Thoret S, Pernod G, Bosson JL; GIHP-NACO Study Group. Management of Severe Bleeding in Patients Treated with

Direct Oral Anticoagulants: An Observational Registry Analysis. *Anesthesiology*. 2017 Apr 14. (investigateur)

- Blanié A, Figueiredo S, Lorre G, Benhamou D. Transient neurologic symptoms after intrathecal injection of ropivacaine through a dural tap during an attempted epidural for labour pain relief. *Anaesth Crit Care Pain Med*. 2017 May 22.
- Blanié A, Roulleau P, Mengelle C, Benhamou D. Comparison of learning outcomes between learning roles (spectator and actor) during an immersive simulation. *Anaesth Crit Care Pain Med*. 2017 Aug;36(4):243-244
- Blanié A, Le Guen M. Debriefing values in high-fidelity simulation. *Anaesth Crit Care Pain Med*. 2017 Aug;36(4):201-202.
- Blanié A, Gorse S, Roulleau P, Figueiredo S, Benhamou D. Impact of learners' role (active participant-observer or observer only) on learning outcomes during high-fidelity simulation sessions in anaesthesia: A single center, prospective and randomised study. *Anaesth Crit Care Pain Med*. 2018 Jan 5
- Beaussier M, Parc Y, Guehot J, Cachanado M, Rousseau A, Lescot T; CATCH study investigators. Ropivacaine Preperitoneal Wound Infusion for Pain Relief and Prevention of Incisional Hyperalgesia after Laparoscopic Colorectal Surgery. *Colorectal Dis*. 2018 Jun;20(6):509-519. (investigateur)
- Fessenmeyer C, Ruscio L, Blanié A, De Montblanc J, Benhamou D. Epidural Analgesia after Colorectal Surgery Within an Enhanced Recovery Program. *J Am Coll Surg*. 2018 Mar;226(3):337-338. doi: 10.1016/j.jamcollsurg.2018.01.001.
- Beloeil H, Albaladejo P, Sion A, Durand M, Martinez V, Lasocki S, Futier E, Verzili D, Minville V, Fessenmeyer C, Belbachir A, Aubrun F, Renault A, Bellissant E; OCTOPUS group. Multicentre, prospective, double-blind, randomised controlled clinical trial comparing different non-opioid analgesic combinations with morphine for postoperative analgesia: the OCTOPUS study. *Br J Anaesth*. 2018 Dec 29. pii: S0007-0912(18)31298-4. doi: 10.1016/j.bja.2018.10.058. (Investigateur)
- Blanié A, Benhamou D. Do observers and active participants learn similarly during high fidelity sessions? *Anaesth Crit Care Pain Med*. 2019 Feb 25. pii: S2352-5568(19)30073-6. doi: 10.1016/j.accpm.2019.02.009.
- Meffert A, Baudrier D, Benhamou D, Mattatia L, Blanié A. Use of cognitive aids in anaesthesia crisis scenarios: Observational analysis from video recording during simulation. *Anaesth Crit Care Pain Med*. 2020 Feb 21. pii: S2352-5568(20)30028-X. doi: 10.1016/j.accpm.2019.12.010.
- Blanié A, Amorim MA, Benhamou D. 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 Med Educ*. 2020 Feb 19;20(1):53. doi: 10.1186/s12909-020-1939-6.
- L'Her E, Geeraerts T, Desclefs JP, Benhamou D, Blanié A, Cerf C, Delmas V, Jourdain M, Lecomte F, Ouane I, Garnier M, Mossadegh C. Simulation-based teaching in critical care, anaesthesia and emergency medicine. *Anaesth Crit Care Pain Med*. 2020 Mar 26. pii: S2352-5568(20)30042-4. doi: 10.1016/j.accpm.2020.03.010.

Communications scientifiques et didactiques

- **Plusieurs posters présentés : à la SFAR (2010 à 2019) ; à la Sofrasims (2015 à 2019)**
- **Communications au congrès de la SFAR :** Enquête sur la connaissance et l'utilisation des aides cognitives de crise en anesthésie réanimation en 2017 (France-Canada). Session CAMR (2018)
- **Communications au congrès SEGAMED :** Mise en place d'un jeu sérieux pour les étudiants en médecine (2014); Mise en place d'un jeu sérieux pour les étudiants infirmiers: LabForGames Warning. Premières évaluations de LabForGames Neuro (2016)
- **Communications au congrès Sofrasims :** Détection de la détérioration d'un patient: exemple d'interprofessionnalité (2016) ; Patient simulé : moi toujours (2018)
- **Communications au congrès MAPAR :** Détection de la détérioration d'un patient hospitalisé. (2016). Aides cognitives de crise (2017); Décision médicale : comment le cerveau raisonne (2019)
- **Communications didactiques à d'autres congrès :**
 - « Simulation et approche à priori ». Congrès JIQHS 2017
 - « Recommandations en simulation des soins critiques ». Colloque des CESU 2019
- **Communications au DU Formation en simulation et Immersion, Université de Lyon :** « Pédagogie Médicale et démarche diagnostique » (2017, 2018, 2019)
- **Chairman :** Débriefing (Sofrasims 2018), session SFAR-Sofrasims (SFAR 2019)
- **Protocoles du MAPAR 2013, 2016 et 2019.** Comité de lecture et rédaction de chapitres et-fiches urgences
- **Participation à la rédaction des textes du CAMR et aides cognitives de crise depuis 2017 :** aides cognitives de crise en anesthésie réanimation ; consultations d'anesthésie délocalisée, itératives ou en télémedecine...

Liste des communications scientifiques et didactiques :

- Odri A et al. Peut-on utiliser la succinylcholine en réanimation ? Résultats préliminaires. Abstract + Communication orale. SFAR 2008.
- Blanié A et al. A628 - Cortex and deep grey matter have different sensitivities to hypoxia-hypotension and traumatic brain injury: a combined PtIO₂ and microdialysis study in rats. Poster à Bruxelles. SFAR 2010.
- Blanié A, et al. Is it safe to use succinylcholine in the intensive care unit? Prospective and observational study. Abstract + communication orale. SFAR 2010.
- Blanié A et al. Durée de la fibrinolyse post opératoire après prothèse totale de hanche et genou : étude de suivi biologique. Poster + communication orale. SFAR 2011.
- Blanié A, Vigué B, Mégret F, Ouatarra A. QFP : La Succinylcholine en réanimation. oct 2012. <http://www.sfar.org/article/943/la-succinylcholine-en-reanimation-qfp-2012>
- Protocoles du MAPAR edition 2013: Comité de lecture des protocoles section Anesthésie et rédaction de plusieurs chapitres.
- Blanié A, Roulleau P, Benhamou D. Evaluation de la formation AFGSU2 : résultats préliminaires. Abstract + communication orale. SFAR 2013

- Blanié A, Roulleau P, Benhamou D. Mise en place d'un jeu sérieux pour les étudiants en médecine : LabForGames. Présentation orale SEGAMED 2014. Abstract et communication chairman. Sofrasims 2015.
- Blanié A. L'apprentissage par la simulation dans la formation des professionnels de santé. Communication didactique. SimDef 2015
- Blanié A, et al. Comparaison de l'apprentissage des connaissances médicales en fonction du rôle de l'apprenant (acteur ou spectateur) lors de simulation haute fidélité : étude préliminaire. Abstract + communication orale à la SFAR2015 et au colloque francophone de simulation en Santé 2015.
- Blanié A, Benhamou D. Evaluation et niveau de rétention après la formation AFGSU2. Abstract + communication orale à la SFAR 2015.
- Protocoles du MAPAR edition 2016. Comité de lecture des protocoles section Anesthésie et rédaction des chapitres dont Fiches urgences.
- Blanié A, Benhamou D. Détection de la deterioration d'un patient: exemple d'interprofessionalité. Communication didactique. Congrès national de la Sofrasims 2016
- Blanié A. Détection de la deterioration d'un patient hospitalisé. Communication didactique. Congrès du MAPAR 2016.
- Blanié A, et al. Evaluation de la connaissance des internes d'anesthésie-réanimation sur la gestion des risques et compétences non techniques. R541- Abstract + Ecommunication SFAR 2016
- Blanié A. Cas Cliniques: gestion des urgences chirurgicales chez un patient sous AOD. Communication didactique et modérateur au symposium Boehringer. SFAR 2016
- Blanié A, Benhamou D. La réponse à l'appel à projets: levier d'innovation dans les formations des professionnels de santé. Communication au colloque paramédicaux ARS 2016
- LabForSIMS centre de simulation en Santé. Journée initiative pédagogique Université Paris Saclay. Novembre 2016
- Blanié et al. Mise en place d'un jeu sérieux pour les étudiants infirmiers: LabForGames Warning. Premières évaluations de LabForGames Neuro. Communications didactiques à SEGAMED 2016.
- Blanié A. Compétences cognitives en simulation et pédagogie. Communication didactique. Congrès DAQUAR 2017
- Blanié A. Aides cognitives de crise. Communication didactique. Congrès MAPAR 2017
- Blanié A. Serious games et raisonnement clinique. Communication didactique. Congrès Sofrasims 2017
- Blanié A. Aides cognitives de crise. Communication didactique. Congrès Sofrasims 2017
- Blanié A. La simulation en santé. Communication didactique. Master 2 science chirurgicale, UE Nouvelles technologies interventionnelles- Faculté de Médecine Paris Sud 2017
- Gestion de crise – Workshop Simulation haute fidélité. Formateur. ESA genève 2017
- Blanié A. « Pédagogie Médicale et démarche diagnostique » ; Communication didactique. DU Formation en simulation et Immersion, Université de Lyon en 2017
- Blanié A. Simulation et approche à priori. Communication didactique. Congrès JIQHS 2017
- Blanié A. Patient simulé : moi toujours. Communication didactique. Congrès Sofrasims 2018
- Débriefing. Chairman. Congrès Sofrasims 2018
- Mohamed-Soilihi A, Blanié B, Bech B, Boithias C, Brouquet A, de Chalendar G, Cauchon L, Girard A, Jourdain G, Benhamou D. LabForSIMS 2 : Simulation ultra-haute fidélité et jeu sérieux de 2nde génération - Bilan à 18 mois d'avancée du projet. Présentation au congrès de la SoFraSimS, Caen 4-6 avril 2018.
- Suet G, Blanié A, Benhamou D. Impact sur l'apprentissage des compétences techniques de l'utilisation d'une grille d'observation chez des internes d'anesthésie-réanimation observateurs en séance de simulation : étude randomisée. Présentation au congrès de la SoFraSimS, Caen 4-6 avril 2018 et congrès SFAR 2018

- Vermersch C, Perrot C, Blanié A, Benhamou D. Formation initiale interprofessionnelle par la simulation: le parcours de soin périnatal de l'éclampsie. Présentation au congrès de la SoFraSimS, Caen 4-6 avril 2018.
- Gestion de crise – Workshop Simulation haute fidélité. Formateur. ESA copenhagen en 2018
- Blanié A et Kureck M. Enquête sur la connaissance et l'utilisation des aides cognitives de crise en anesthésie réanimation en 2017: en France et au Canada. Communication didactique. Session CAMR. Congrès SFAR 2018
- Blanié A. « Pédagogie Médicale et démarche diagnostique » ; Communication didactique. DU Formation en simulation et Immersion, Université de Lyon en 2018
- Protocoles du MAPAR edition 2019. Comité de lecture des protocoles section Anesthésie et rédaction des chapitres dont Fiches urgences.
- Blanié A, Romero C, Girard A, Boithias C, Jourdain G, Mohamed-Soilihi A, Benhamou D. Mal des simulateurs lors d'une application médicale de réalité virtuelle. Présentation au congrès de la SoFraSimS, Strasbourg 15-17 mai 2019.
- Dupont R, Prokopetc K, Boithias C, Jourdain G, Blanié A, Bech B (2), Mohamed-soilihi A, Benhamou D. Réalité mixte pour la formation à la réanimation néonatale. Présentation au congrès de la SoFraSimS, Strasbourg 15-17 mai 2019.
- Laleye F, Blanié A, Brouquet A, Bech B, Mohamed-Soilihi A, Benhamou D, De Chalendar G. Agents conversationnels pour la formation au diagnostic medical. Présentation au congrès de la SoFraSimS, Strasbourg 15-17 mai 2019.
- Blanie A, G Arcile, P Roulleau, Benhamou D. Impact sur l'apprentissage des compétences non techniques de l'utilisation d'une grille d'observation non technique chez des internes d'anesthésie-réanimation observateurs en séance de simulation : étude randomisée. Présentation au congrès de la SoFraSimS, Strasbourg 15-17 mai 2019.
- Blanie A, Amorim MA, Benhamou D. Comparaison de la valeur pédagogique d'un enseignement (par jeux sérieux versus traditionnel) testant l'apprentissage du raisonnement clinique lors de la détection de la dégradation de l'état clinique d'un patient chez les étudiants en soins infirmiers. Présentation au congrès de la SoFraSimS, Strasbourg 15-17 mai 2019.
- Gestion de crise – workshop Simulation haute fidélité. Formateur. ESA Vienne en 2019
- Laleye F, Blanié A Brouquet A Benhamou D, de Chalendar G. Hybridation d'un agent conversationnel avec des plongements lexicaux pour la formation au diagnostic médical. TALN 2019
- Blanié A: « recommandations en simulation des soins critiques » Communication au colloque des CESU (2019)

Communication - Presse

- Reportage audio sur Radio Classique dans le cadre des portes ouvertes de l'AP-HP (2015)
- Reportage sur le LabForSIMS dans le magazine de la santé. sur France 5. (2015) http://www.allodocteurs.fr/se-soigner/professions-medicales/les-futurs-medecins-formes-dans-les-conditions-du-reel_17656.html
- Interview pour Article en ligne du Parisien paru à l'occasion des portes ouvertes de l'APHP et du centre de Simulation LabForSIMS (2016). <http://www.leparisien.fr/le-kremlin-bicetre-94270/le-kremlin-bicetre-decouvrez-le-centre-de-simulation-de-medecine-27-05-2016-5834761.php>

- Reportage au JT de M6 sur les Serious Game (2016)
- Interview pour un article dans le cadre de la formation aux gestes élémentaires avant la première garde des étudiants de médecine.
http://www.lequotidiendumedecin.fr/actualites/article/2016/09/02/paris-sud-les-etudiants-sexerceront-sur-des-mannequins-avant-leur-premiere-garde_823736
Reportage AFP -

Titre : Évaluation expérimentale du raisonnement clinique dans le cadre des jeux sérieux pour la formation des professionnels de santé

Mots clés : Jeux sérieux, simulation, formation, professionnels de santé, raisonnement clinique

Résumé : L'amélioration du raisonnement clinique (RC) est un enjeu essentiel pour la Médecine de demain car il est établi que son utilisation imparfaite conduit à des résultats de soins insuffisants. Le RC est un processus cognitif complexe. Cette activité intellectuelle synthétise l'information obtenue à partir de la situation clinique et l'utilise pour faire une analyse diagnostique et prendre une décision de prise en charge du patient en intégrant les connaissances et expériences antérieures. La formation à cette compétence est donc essentielle. Pour améliorer le raisonnement, une connaissance des mécanismes qui le constituent est nécessaire et une revue de ces mécanismes constitue la partie initiale de cette thèse.

La formation des professionnels de santé par la simulation se généralise avec pour objectif « jamais la première fois sur le patient ». En plein essor, le jeu sérieux (JS) représente un outil pédagogique intéressant. Une revue de la littérature sur l'efficacité des JS et plus particulièrement dans le cadre du RC est également incluse dans la partie initiale de cette thèse. Ainsi, les JS sont efficaces et peuvent, entre autres, cibler certaines compétences dont le RC. Cependant, la plupart des études sur le RC dans le cadre des JS, restent subjectives avec des évaluations qualitatives ou des autoévaluations des apprenants ou concernent uniquement le résultat (prise de décision). Par conséquent, la valeur éducative et les modalités des JS dans la formation du RC restent à approfondir.

Le JS *LabForGames Warning* a été développé dans le centre de simulation LabForSIMS pour les étudiants infirmiers et cible la détection de l'aggravation d'un patient et la communication. Cette thèse a pour objectif de tester un mode d'apprentissage utilisant la simulation par les JS afin d'améliorer le RC chez les professionnels de santé. Une 1^{ère} étude a évalué la validité du JS *LabForGames Warning* selon le cadre théorique proposé par Messick. Elle a démontré que les scores et le temps de jeu ne pouvaient pas différencier le niveau des compétences cliniques des infirmiers.

Cependant, les preuves de validité étaient obtenues pour le contenu, le processus de réponse et la structure interne. Même si cette version du jeu ne peut donc pas être utilisée pour une évaluation sommative des étudiants, notre étude montre que ce JS est bien accepté par les étudiants et qu'il peut être utilisé pour la formation au sein d'un programme éducatif. Une seconde étude a évalué l'efficacité de 2 modalités pédagogiques sur l'apprentissage du RC à la détection de l'aggravation clinique d'un patient en comparant un groupe d'étudiants infirmiers formé par simulation avec *LabForGames Warning* par rapport à un groupe formé par un enseignement traditionnel. Le RC a été évalué par les tests de concordance de script immédiatement et 1 mois après. Cette étude multicentrique randomisée a inclus 146 étudiants infirmiers volontaires. Aucune différence significative n'a été observée sur le RC entre la formation par la simulation avec JS et l'enseignement traditionnel. Cependant, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation.

En conclusion, nous avons tout d'abord confirmé la validité du JS *LabForGames Warning* en tant qu'outil pédagogique à visée formative et non sommative. Puis, bien qu'aucune différence d'apprentissage du RC n'ait été observée entre la formation par la simulation avec JS et l'enseignement traditionnel, la satisfaction et la motivation étaient meilleures avec l'enseignement par simulation avec le jeu. Les études sont à poursuivre pour préciser les modalités et stratégies pédagogiques des JS dans la formation des professionnels de santé, comme par exemple la place du débriefing, le rôle de la motivation. En effet, en plein essor, les développements technologiques telle que l'intelligence artificielle vont transformer la formation au RC ainsi que les outils pédagogiques disponibles dans les années à venir.

Title : Experimental Evaluation of Clinical Reasoning in Serious Games for the Education of Health Professionals

Keywords : Serious games, simulation, training, healthcare professionals, clinical reasoning

Abstract : Improvement of clinical reasoning (CR) is a key issue for the future of medicine because it has been established that imperfect reasoning leads to insufficient care results. CR is a complex cognitive process which synthesizes information obtained from the clinical situation, then uses it to make a diagnostic analysis and take a decision on patient management by integrating previous knowledge and experience. Training for this skill is therefore essential. To improve reasoning, knowledge of the mechanisms which build it up is necessary and a review of these mechanisms constitutes the initial part of this thesis.

Training of healthcare professionals through simulation is becoming widespread with the objective of "never the first time on a patient". In rapid expansion, the use of serious games (SG) represents an interesting pedagogical tool. A review of the literature on the effectiveness of SG and more particularly in the context of CR is also included in the initial part of this thesis. Thus, SG is effective and may, among other things, target certain skills, including CR. However, most studies dealing with CR by using SG include qualitative assessments or self-assessments of learners or focus only on the outcome (decision making). Therefore, the educational value and modalities of SG in the training of CR of health professionals remain to be further explored.

The SG *LabForGames Warning* was developed in the LabForSIMS simulation center for nursing students and targets detection of patient deterioration and the ensuing communication. The objective of this thesis is to test a learning mode using SG simulation in order to improve CR in healthcare professionals.

The first study evaluated the validity of the SG *LabForGames Warning* according to the Messick's Framework. This study showed that scores and playing time could not differentiate the level of clinical skills of nurses. However, evidence of validity was obtained for content, response process and internal structure. Although this version of the game cannot therefore be used for summative evaluation of students, our study shows that this SG is well accepted by students and can be used for training within an educational program.

A second study evaluated the effectiveness of 2 teaching modalities on learning CR for the detection of clinical patient deterioration by comparing a group of nursing students trained by simulation with *LabForGames Warning* compared to a group trained by traditional teaching. CR was assessed by script concordance tests immediately and 1 month later. This randomized multicentre study included 146 volunteer nursing students. No significant difference was observed in CR change between simulation training with SG and traditional teaching. However, satisfaction and motivation were better with simulation instruction.

In conclusion, we have confirmed the validity of SG *LabForGames Warning* as an educational tool with formative and not summative aims. Then, although no difference in learning about CR was observed between simulation training with SG and traditional teaching, satisfaction and motivation were better with simulation teaching with the game. Further studies are needed to clarify the modalities and pedagogical strategies of SG in the training of healthcare professionals, such as the place of debriefing and the role of motivation. Moreover, technological developments such as artificial intelligence might transform CR training and the available pedagogical tools in the coming years