

Développement d'un outil d'évaluation des compétences des étudiants en médecine d'urgence : une étude nationale par la méthode Delphi

The Development of a Competencies Assessment's Tool for Emergency Medical Students and Residents: a Nationwide Delphi Study

A.-L. Philippon · P. Hausfater · E. Tribby · Y. Freund

Reçu le 11 juillet 2019 ; accepté le 21 octobre 2019
© SFMU et Lavoisier SAS 2019

Résumé Objectif : Ce travail s'inscrit dans une démarche d'ingénierie des compétences visant à évaluer par la simulation les étudiants et les internes en médecine d'urgence (MU). L'objectif de l'étude est de développer un score d'évaluation par la simulation pour trois familles de situation (FS).

Matériels et méthodes : Le choix des trois FS a été fait après un classement selon leur importance, réalisé par les coordonnateurs régionaux du diplôme d'études spécialisé de MU. La seconde étape consistait en une méthode Delphi, pour élaborer une grille d'évaluation des étudiants sur ces FS. Parmi les 86 experts sollicités, l'objectif était d'avoir 20 experts pour chaque FS, qui devraient noter des items en fonction de leur pertinence à évaluer, sur une échelle de Likert pour le premier et le deuxième tour puis sur un score binaire au tour final. Le consensus a priori était de 75 % pour les deux premiers tours puis de 85 % pour le dernier tour.

Résultats : Les trois FS identifiées sont l'arrêt cardiorespiratoire, la détresse respiratoire aiguë et le coma non traumatique. Cinquante et un experts ont accepté de participer (59 %). Le taux de participation a varié de 80 à 95 % pour chaque tour, et les experts ont identifié 20 catégories composant un score d'évaluation pour chaque situation.

Conclusion : Grâce à une méthode de consensus de groupe,

nous avons pu déterminer les items indispensables à l'évaluation en simulation de trois situations cliniques graves en MU.

Mots clés Pédagogie médicale · Simulation · Évaluation · Médecine d'urgence · Delphi

Abstract Aim: Our work is part of a competence-based assessment approach. Our aim was to determine three clinical situations that need a simulation-based-assessment for medical students and residents and then, to develop a score for each situation.

Material and methods: We have conducted two consensus groups' methods. First, we have asked all French residency program directors in emergency medicine to determine the clinical situations that we had to score. The second step consisted in the use of a Delphi technique, in three rounds, to achieve consensus and to select items that had to be part of the assessment score. Among the 86 solicited experts, we aimed to have 20 experts for each of the three scores. Within the rounds 1 and 2, experts were asked to note items on a Likert scale, and for the last round on a binary scale. For rounds 1 and 2, consensus rate of 75% was agreed upon. It was 85% for round 3.

Results: The directors identified those clinical situations: cardiac arrest, acute respiratory failure and non-traumatic coma. Response rates for the Delphi were 59% for the overall participation, and then varied from 80 to 95% in each round, for each score. We have found 20 categories for each situation, and made scores with those categories.

Conclusion: With the help of consensus groups' methods, we have found three clinical situations important to assess by simulation in emergency medicine, and developed a score for each situation, made of 20 categories.

Keywords MeSH: Medical education · Simulation training · Educational measurement · Emergency medicine · Delphi Technique

A.-L. Philippon (✉) · P. Hausfater · Y. Freund
Service d'accueil des urgences,
groupe hospitalier Pitié-Salpêtrière Charles-Foix,
Assistance publique-Hôpitaux de Paris (AP-HP),
F-75013 Paris, France
e-mail : annelaurephi@gmail.com

A.-L. Philippon · E. Tribby
Laboratoire interuniversitaire des sciences de l'éducation
et de la communication (LISEC), EA 2310,
université de Strasbourg, France

P. Hausfater · Y. Freund
Faculté de médecine, Sorbonne Université,
F-75013 Paris, France

Introduction

Un des enjeux actuels des disciplines universitaires et notamment médicales est d'évaluer et de certifier les étudiants selon le modèle de l'approche par compétences [1,2]. Selon la définition de Jouquan et Parent, une compétence est « *un savoir-agir complexe et singulier qui mobilise, grâce à des capacités de nature diverse (cognitive, réflexive, métacognitive, opérative, sociale), un ensemble de ressources complémentaires, élaborées à partir de savoirs multiples et organisées en schémas opératoires, pour traiter de façon adéquate des problèmes à l'intérieur de familles de situations professionnelles, définies au regard de rôles, de contextes et de contraintes spécifiques* » [3]. Ainsi, le caractère complexe de la compétence met en évidence des notions que les formateurs en médecine d'urgence (MU) doivent avoir à l'esprit : la compétence consiste en une mobilisation efficace de savoirs, d'habiletés, en s'appuyant sur des ressources (acquis universitaires, expériences, centre d'intérêts mais également les pairs, enseignants, collègues) pour résoudre des problèmes au sein de familles de situation (FS) [4,5]. La compétence devient donc totalement indissociable des situations professionnelles que rencontrera le futur médecin urgentiste.

La notion de FS devient alors primordiale. Elle permet de regrouper différentes situations semblables par leur présentation ou leur nature, mais qui diffèrent par leur contexte [4]. En MU, une FS serait par exemple une présentation clinique, la détresse respiratoire aiguë, qui présente de nombreux points communs ou invariants dans la prise en charge mais que l'étudiant compétent saura appréhender en fonction des différentes étiologies, contextes (intra- ou extrahospitalier), en fonction du patient lui-même, de ses collègues, etc. Ainsi, l'intérêt de raisonner avec les FS permet d'identifier les situations professionnelles propres aux urgences puis de pouvoir identifier des points communs (ou invariants) à chaque situation pour ensuite de faire varier le contexte et donc les situations d'apprentissage [4,6].

En France, sept compétences médicales dites génériques ont été identifiées et leur place réaffirmée en 2013, lors de la réforme du deuxième cycle [7]. Il s'agit en particulier, des compétences de clinicien, de communicateur et de membre d'une équipe soignante pluriprofessionnelle [7,8]. Leur évaluation devient complexe puisqu'il s'agit non plus d'évaluer des connaissances théoriques ou pratiques pures, mais ce que l'on nomme les « composantes des compétences », à savoir, la capacité qu'aura l'étudiant à mobiliser ses ressources, à raisonner, à communiquer avec une équipe, tout en démontrant sa maîtrise pour la réalisation de gestes techniques. Les outils tels que les questions à choix multiples actuellement utilisés pour évaluer les étudiants ont montré leurs limites comme outil d'évaluation des compétences d'apprenants pour des situations complexes [9,10].

La simulation en santé offre une opportunité pour répondre à l'enjeu d'évaluations plus « authentiques » car permettant une mise en situation des étudiants d'autant qu'elle commence à apporter des preuves de son efficacité sur l'amélioration de la performance individuelle et sur l'apprentissage de certains gestes et de certains processus de soins. Les preuves restent toutefois encore peu nombreuses quant aux niveaux plus élevés de l'évaluation des effets produits par un dispositif de formation. Il s'agit en particulier des changements de pratique professionnelle des personnes formées et d'un impact positif direct chez les patients [11–13]. Pour réaliser une évaluation valide, il faut disposer de scores reproductibles et fiables. De tels scores existent, mais ils évaluent soit des gestes techniques, soit des capacités de travail en équipe et ne sont en général pas destinés aux étudiants en formation initiale [14,15]. De plus, la validité d'une évaluation doit être envisagée avec une approche plus globale, et il faut chercher ce que Downing appelle les différentes « sources » de validité d'une évaluation. Il en cite au moins cinq, qui sont le plus couramment utilisées et qui sont : le contenu de l'évaluation représenté par sa signification, sa cohérence avec les objectifs pédagogiques ; le processus de réponse utilisé par les étudiants ; la structure interne du score, représentée par ses caractéristiques psychométriques ; les relations entre les résultats à l'évaluation et les résultats obtenus à une autre évaluation d'un contenu identique et enfin, les « conséquences » des résultats pour les étudiants, à savoir l'impact sur l'apprentissage ultérieur, sur leur pratique... [16].

La problématique de l'évaluation est d'autant plus importante à l'heure de la naissance de la spécialité de MU, qui nécessite un enseignement de qualité avec un haut niveau d'exigence au sein d'un cursus de diplôme d'études spécialisées (DES) dédié à la formation des futurs urgentistes [17,18]. Mais cette problématique concerne également les étudiants de deuxième cycle qui, pendant leur formation initiale, doivent acquérir des compétences en MU, afin d'initier la prise en charge des situations de défaillances vitales qu'ils rencontreront lors de leur internat et dans leur exercice futur.

Face à ces enjeux importants, l'objectif de ce travail était de déterminer trois principales FS en MU qui devraient faire l'objet d'une évaluation des étudiants et d'élaborer pour chacune d'elle une grille de notation afin d'évaluer les étudiants au cours de séances de simulation en santé.

Matériels et méthodes

L'étude a été menée en deux temps, de novembre 2017 à janvier 2018 puis de mars à octobre 2018. Pour ces deux périodes, un groupe d'experts a été constitué afin de déterminer les FS, puis pour la détermination par méthode Delphi

d'un score pour chaque FS. L'avis d'un comité d'éthique n'était pas requis pour ce type de recherche [19].

Lors de la seconde période, la méthode Delphi utilisée a respecté les cinq principes fondamentaux d'une consultation d'expert qui sont : le caractère anonyme des participants, la répétition des consultations, faite de feed-back sur les réponses des experts, des interactions entre les experts contrôlées par les investigateurs et une analyse statistique des réponses [20,21]. De plus, ont été respectées les différentes étapes recommandées : la réflexion autour de la problématique de recherche puis l'analyse de la littérature, le choix des experts, la rédaction des questions et du questionnaire, les différents tours de consultation d'experts, dont chacun est suivi d'un retour sur les réponses de chacun, la définition du consensus et, enfin, l'analyse des résultats [22].

Le panel d'expert a été composé différemment lors des deux périodes de l'étude. Les experts de la première période étaient tous les coordonnateurs régionaux du DES de MU (DES-MU). Pour la méthode Delphi, les critères de sélection des experts étaient les suivants : ils devaient être francophones, avoir des responsabilités dans l'enseignement de la MU et/ou avoir une expertise acquise par une longue expérience d'enseignement (> 5 ans). Par ailleurs, il fallait également avoir des notions de l'enseignement par simulation ainsi que de ses spécificités. Les experts ont été recrutés par contact e-mail. Les critères de leur sélection leur ont été soumis afin de vérifier que leur profil correspondait bien. Ils pouvaient également suggérer des experts qui, selon eux, correspondaient aux critères. Ils ne savaient pas qui avait été contacté ni qui participait finalement au Delphi. Notre objectif était d'avoir 20 experts par FS.

Le premier questionnaire a été envoyé aux 29 coordonnateurs de DES-MU régionaux, à qui nous avons demandé de choisir puis de classer par ordre d'importance dix items correspondant à des FS qui, selon eux, devraient faire l'objet d'une évaluation des étudiants en médecine et internes en MU par la simulation. Les dix items étaient à choisir parmi 36 items figurant dans le référentiel de formation à destination des étudiants en médecine « Urgences et défaillances viscérales aiguës » [23]. Suite à cette consultation, les trois FS ayant obtenu le plus de points ont été sélectionnées.

Le second questionnaire a été envoyé aux experts sélectionnés qui ont accepté de participer à l'étude. Il avait pour but d'établir un score à partir d'un grand nombre d'items choisis par les investigateurs selon le référentiel de formation, mais également selon les différentes recommandations en vigueur pour chaque FS. Ces items pouvaient être d'ordre cognitif (par exemple, rechercher les facteurs de risque cardiovasculaire), opératif (réaliser un massage cardiaque externe de qualité) ou encore social et psychoaffectif (appeler la famille, se présenter au patient, communiquer avec l'équipe, etc.).

Lors du premier tour, nous avons demandé aux experts de noter les items qui leur paraissaient pertinents à la constitution d'un score qui permettrait d'évaluer un étudiant avec la simulation. Il était précisé que le score devrait pouvoir évaluer des étudiants en formation initiale, mais ayant différents niveaux de compétence, de la fin du deuxième cycle jusqu'à la phase de consolidation du troisième cycle. Les experts devaient noter chaque item grâce à une échelle de Likert, allant de 1 (pas du tout pertinent) à 6 (tout à fait pertinent). Il était également demandé aux experts d'ajouter des items non cités qui leur paraissaient pertinents. Le deuxième tour s'est déroulé de manière identique, après retrait ou ajout de certains items, et renvoi aux experts des items retenus et des commentaires de chacun d'entre eux, de manière anonyme. Certains items pouvaient également être reformulés et fusionnés suite à l'analyse des commentaires. Le troisième tour était différent : tous les items retenus au tour 1 et 2 étaient soumis à évaluation, afin que les experts aient un aperçu global du score. La notation des items devenait binaire : 1 pour item indispensable, 0 pour item à ne pas retenir dans le score final (Annexe A).

Définition des consensus, arrêt du Delphi et analyse statistique

Pour le Delphi, il avait été déterminé à l'avance qu'il serait composé de trois tours, et que l'objectif était d'obtenir des scores composés de 20 catégories. Chaque catégorie correspondrait à un élément d'évaluation de l'étudiant et pourrait être composée d'un seul item ou en regrouper plusieurs. Par exemple, une catégorie peut regrouper deux items tels que mesurer la fréquence respiratoire et la saturation en oxygène. Pour le premier et le deuxième tour, le consensus positif était atteint lorsque plus de 75 % des experts estimaient l'item pertinent, soit un premier percentile (Q1) supérieur à 5 : l'item était conservé pour le futur score. Le consensus négatif était défini par 75 % des experts qui estimaient l'item non pertinent, soit un troisième percentile (Q3) inférieur à 5. Tous les items ne correspondant pas à ces définitions étaient remis au jugement des experts. Pour le troisième tour, comme le score était binaire, le consensus était défini par un accord d'expert supérieur à 85 %. À chaque tour, les remarques des experts ont été également prises en compte et analysées pour modifier la formulation de certains items, les adapter à la simulation et pour en regrouper certains d'entre eux.

Résultats

Parmi les 29 coordonnateurs régionaux sollicités, 21 (72 %) ont participé et permis d'identifier les dix FS à évaluer, les

trois premières étant : l'arrêt cardiorespiratoire, la détresse respiratoire aiguë et le coma non traumatique.

Nous avons sollicité 58 experts francophones pour participer au Delphi qui, à leur tour, nous ont indiqué 28 experts qui répondaient aux critères d'inclusion. Sur les 86 experts sollicités, 51 (59%) ont accepté de participer au Delphi (Tableau 1). Afin d'avoir 20 experts pour la constitution de chaque score, nous avons demandé à cinq experts de participer à l'établissement de plusieurs scores. Il s'agissait d'experts ayant une réflexion sur l'évaluation des étudiants par la simulation. Le taux de réponse a varié entre 80 et 95 % (Tableau 2).

Pour chaque FS, ont été obtenus au troisième tour 24, 26 ou 30 items que nous avons regroupés en 20 catégories pour établir un score sur 20 (Fig. 1). Notre score d'évaluation n'étant pas un score binaire, cela permet de regrouper pour une seule catégorie, plusieurs items. Cette « fusion » d'items a pu se faire dans de nombreux cas grâce à l'analyse des remarques des experts.

Tableau 1 Description des 51 experts qui ont participé au Delphi	
	Nombre (%)
Total participants	51
Fonction - Spécialité	
Universitaire MU	17 (33)
Universitaire AR	3 (6)
Médecin urgentiste	26 (51)
Médecin AR	3 (6)
Autre	2 (4)
Responsable DES-MU/DESC	9 (18)
Chef de service	8 (16)
Lieu d'exercice	
France	
Auvergne-Rhône-Alpes	6 (23)
Bourgogne-Franche-Comté	3 (6)
Centre-Val de Loire	1 (2)
Grand Est	2 (4)
Hauts-de-France	2 (4)
Île-de-France	23 (45)
Nouvelle Aquitaine	4 (8)
Occitanie	4 (8)
Pays de la Loire	1 (2)
Suisse	2 (4)
Belgique	3 (6)
Type d'hôpital	
Centre hospitalier universitaire	49 (96)
Centre hospitalier	2 (4)

MU : médecine d'urgence ; AR : anesthésie-réanimation ; DES-MU : diplôme d'études spécialisées en médecine d'urgence ; DESC : diplôme d'études spécialisées complémentaires

Tableau 2 Taux de réponse des experts pour chaque famille de situation (FS) et à chaque tour de Delphi

Famille de situation	Tour 1	Tour 2	Tour 3
ACR	19 (95)	18 (90)	18 (90)
DRA	17 (85)	17 (85)	16 (80)
Coma non traumatique	17 (85)	17 (85)	17 (85)

ACR : arrêt cardiorespiratoire ; DRA : détresse respiratoire aiguë. Les résultats sont exprimés en nombre (%)

L'analyse des commentaires des experts a permis d'améliorer chaque score et a également permis de relever des remarques utiles à la mise en pratique de l'évaluation. Les principales problématiques soulevées par les experts étaient au nombre de trois : problèmes de contextualisation, discussions sur les limites et les pratiques de la simulation et questionnement sur les méthodes de l'évaluation.

La principale difficulté exprimée par les experts était l'absence de contextualisation, et ce, pour deux raisons. La première est que le public évalué est de niveau variable : d'étudiant de deuxième cycle à interne en phase de consolidation, et avec donc des objectifs différents en termes de niveau de développement des compétences à acquérir. L'objectif du score étant de pouvoir les discriminer, chaque item a été discuté et remis ou retiré dans ce sens. Les items qui ont été le plus discutés étaient ceux en rapport avec l'intubation oro-trachéale (IOT) et les compétences requises variables pour chaque catégorie d'étudiant. La seconde difficulté de contextualisation résidait dans le concept même de FS et dans le fait que les situations sont assez variées, d'étiologie variable, ce qui a perturbé les experts, habitués à raisonner à partir d'un cas clinique : contexte précis, problématique précise, étiologie précise...

Le deuxième point qui a été source de désaccord parmi les experts est en lien avec la fidélité du mannequin. Ils soulignent en effet la difficulté de calculer un score de Glasgow ou encore de voir des lésions cutanées sur un mannequin. L'absence de fidélité totale était exprimée par quelques experts comme une impossibilité de se servir de cet outil pour évaluer les habiletés cliniques des étudiants. Après analyse, il s'est avéré que la majeure partie des experts a considéré qu'il s'agissait d'un problème inhérent à la simulation, que les étudiants le rencontreraient pendant les phases d'apprentissage et qu'une pratique répétée de la simulation permettrait de limiter les difficultés. Les experts ont donc insisté sur l'absolue nécessité de préciser ces écueils lors du briefing, de recourir à un facilitateur pour répondre aux questions des étudiants à ce propos, (...) mais également d'intégrer dans l'évaluation la phase de débriefing, qui peut permettre, sous certaines conditions, d'explorer et d'évaluer le raisonnement ayant sous-tendu les décisions et actions des étudiants pendant le déroulement du scénario de simulation.

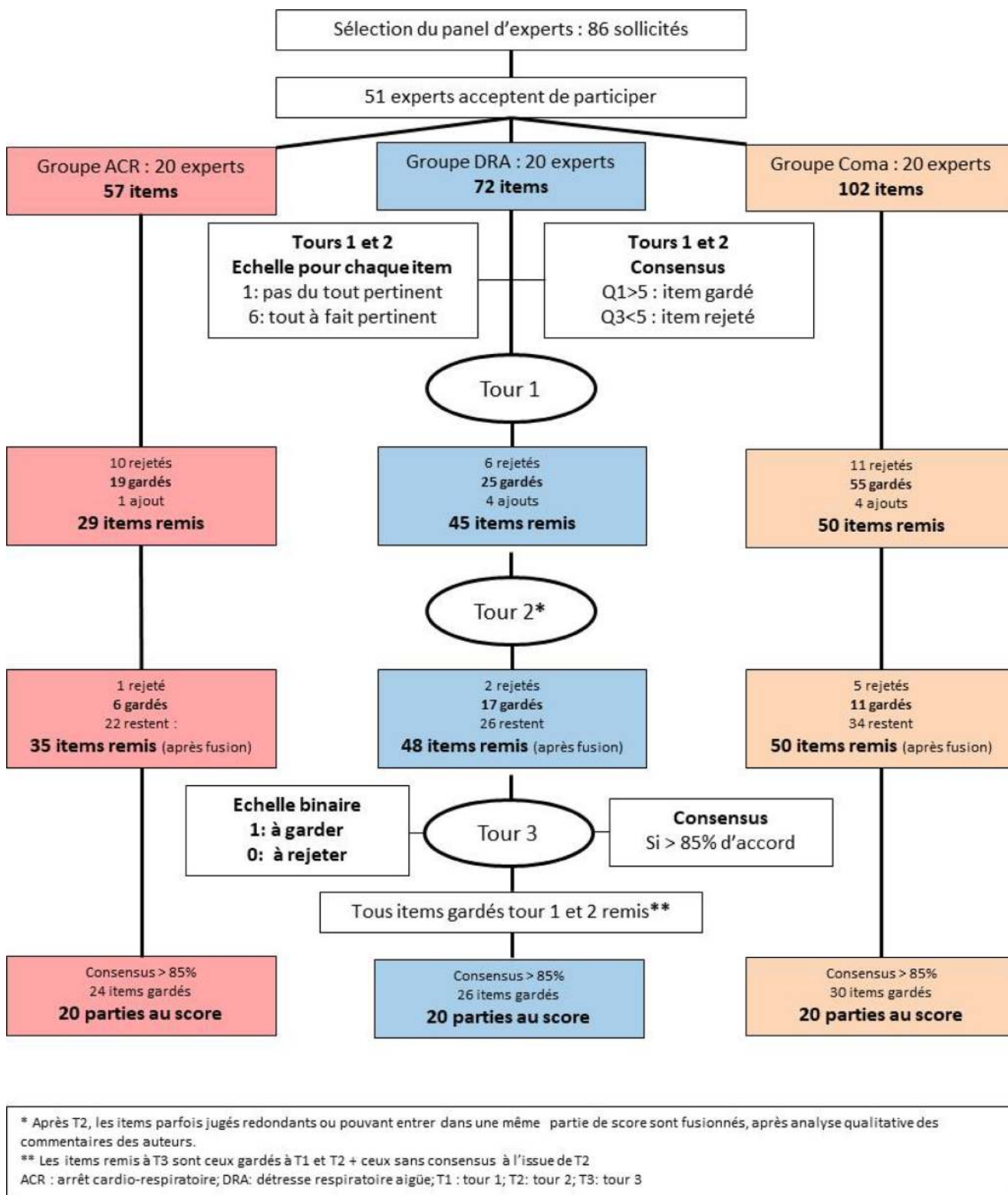


Fig. 1 Résultats du Delphi pour chaque famille de situation

Choisir les items qui feront partie d'un score a permis aux experts de poser la problématique des conditions de l'évaluation, qui n'étaient pas détaillées, volontairement, dans la

présentation de l'étude. Cela a conduit à des propositions sur le déroulement possible de l'évaluation et sur la composition du score (binaire ou pas, avec des items pondérés ou pas).

Discussion

L'utilisation d'une méthode de consensus de groupe, avec un taux de participation correct des experts, a permis de déterminer trois FS importantes à évaluer en MU puis a permis de développer un score composé de 20 catégories pour chaque situation. Les trois FS sélectionnées correspondent toutes à des situations d'urgences vitales que les étudiants en médecine rencontrent peu lors de leurs stages et auxquelles sont trop peu exposés les internes de MU. Cependant, un externe en fin de deuxième cycle doit pouvoir initier leur prise en charge initiale et un interne de MU savoir gérer la situation dans sa globalité, avec de plus en plus d'exigence au fur et à mesure de l'avancement de son cursus. L'arrêt cardiorespiratoire correspond à une situation déjà majoritairement enseignée pour les internes de MU en France [24]. En revanche, lors d'une étude réalisée en 2017, les autres situations retrouvées parmi les plus abordées en simulation étaient l'état de choc, la douleur thoracique, la tachycardie ainsi que le traumatisé grave. Les trois premières thématiques pourraient néanmoins s'intégrer dans nos FS et être évaluées. L'intérêt de travailler dans le cadre des FS est ainsi souligné [4,6].

Cette façon de raisonner n'est pas habituelle en simulation de santé, puisque le plus souvent la FS est mise en pratique au travers d'un cas clinique, qui décrit avec précision un contexte, une étiologie et des objectifs pédagogiques précis [25]. Cela a pu perturber les experts et explique leurs commentaires et leurs difficultés de choix de certains items en rapport avec la contextualisation. Lors de la mise en pratique des scores, il est envisagé que plusieurs cas cliniques seront mis en situation dans le cadre d'une même FS. L'intérêt du score étant de pouvoir évaluer les étudiants dans des contextes différents.

Les experts soulignent dans leurs commentaires les difficultés d'une évaluation par la simulation. Ils relèvent principalement la problématique de fidélité de la situation simulée qui, nous l'avons vu, peut être solutionnée en partie par la tenue d'un briefing de qualité, la présence de facilitateurs et le maintien d'un débriefing après le scénario. Ces trois éléments font partie des bonnes pratiques de la simulation, et il faut s'appuyer sur ces bonnes pratiques pour ne pas réitérer des évaluations sans feedback telles qu'elles sont actuellement réalisées dans le cadre d'une évaluation par QCM par exemple. Le débriefing faisant partie intégrante de la méthode d'apprentissage par simulation, il doit être maintenu et ne doit pas être substitué par au profit d'une évaluation « calquée » sur les examens habituels [26,27]. En effet, l'évaluation par simulation ne peut se résumer à une évaluation de comportements observables, mais bien à une évaluation qui prend en compte le raisonnement de l'étudiant et sa capacité à mobiliser des ressources. Il s'agit d'une autre difficulté des évaluations en situation authentique, mais qui

devrait être prise en compte dans un tel processus, qui vise à améliorer les pratiques évaluatives.

Cette réflexion nous amène à aborder la problématique de l'authenticité d'une évaluation ainsi que la problématique du choix de la simulation pour évaluer les étudiants. En effet, l'approche par compétence sous-entend l'authenticité des activités d'enseignement et d'évaluation afin de pouvoir aborder les situations dans leur complexité et dans leur globalité. Une évaluation « authentique » requiert notamment « une mise en situation » pour « exiger la mobilisation d'une compétence » [3,28]. La situation idéale est le terrain de stage, mais elle présente plusieurs obstacles et notamment celui « d'utiliser » le patient comme outil d'évaluation [29]. Face à cet obstacle de taille, l'utilisation de la simulation, tout en ayant conscience de ses limites, permet de réaliser une évaluation qui tend à s'approcher d'une situation authentique. Ainsi, l'évaluation par la simulation permet de s'approcher d'une évaluation authentique au lit du patient, même si elle présente encore de nombreuses limites, soulignées par les experts : la fidélité du mannequin, la nécessité d'avoir recours à l'interprofessionnalité si on veut s'approcher du réel ou encore des limites inhérentes à la simulation telles que la nécessité de moyens matériels et humains importants. Ces limites représentent les qualités des évaluations actuellement utilisées, écrites, informatisées qui requièrent peu de moyens, mais qui ont une validité écologique très faible [9,10].

Notre étude présente plusieurs limites, inhérentes à la consultation d'experts qui reste une méthode de recueil de consensus. Cependant, le respect des différents critères de qualité a permis d'éviter certains écueils courants avec ces méthodes, tels que le fait de ne pas exposer ses propres opinions ou bien de mener une sélection non pertinente des experts [30]. Nous avons demandé à des experts d'horizons différents de participer et une fois qu'ils ont accepté de faire partie de l'étude, les taux de réponse sont restés satisfaisants, permettant donc une continuité et un suivi des réponses. La majeure partie des experts exerçait en CHU, ce qui est également un biais, mais relève du fait que la simulation se pratique encore pour l'essentiel dans des structures attachées aux universités.

Notre étude constitue une première étape d'un long processus et offre plusieurs perspectives de recherche. La validité des scores obtenus nécessite d'être testée sur des populations différentes d'apprenants : des étudiants en fin de deuxième cycle, des internes de MU en phase socle et en phase de consolidation. Ensuite, il faudrait valider les scores sur des populations différentes : à la fois sur le niveau de compétence pour savoir si les scores sont discriminants, mais également sur les populations différentes. Cette analyse devra être effectuée dans des lieux d'enseignements différents, avec des enseignants et des étudiants qui varient. Enfin, il faudrait également s'assurer de la transférabilité des résultats obtenus lors d'une évaluation qui utiliserait ces scores.

Conclusion

En utilisant une méthode de consensus de groupe type Delphi, nous avons pu déterminer puis établir des scores d'évaluation pour trois détresses vitales en MU : l'arrêt cardiorespiratoire, la détresse respiratoire aiguë et le coma non traumatique. La prochaine étape de notre recherche doit consister en une validation de ces scores.

Liens d'intérêts : les auteurs déclarent ne pas avoir de liens d'intérêts.

Annexe A. Exemples d'items soumis aux experts, pour chaque famille de situation

Détresse respiratoire aiguë

Installer le patient semi-assis (dans certain délai)
Recherche des signes de lutte : faite et communiquée
Se présenter au patient/à sa famille
Annoncer le devenir du patient, en discuter en équipe
Avant de débiter le traitement, explications claires au patient, réassurance

Coma non traumatique

Mesurer la glycémie capillaire
Faire un résumé des éléments fournis par l'anamnèse et l'examen clinique
Intubation : savoir régler le ventilateur
S'assurer de la compréhension des autres membres de l'équipe
Présenter le patient au correspondant : présentation claire et synthétique

Arrêt cardiorespiratoire

Appeler « à l'aide » : autour de soi, prévenir de la survenue de l'arrêt
Débiter le massage cardiaque : évaluer le délai d'initiation du massage
Rechercher une intoxication potentielle
Débiter/organiser le traitement étiologique
Distribuer les rôles

Références

- Epstein RM (2007) Assessment in medical education. *N Engl J Med* 356:387–96
- Holmboe ES, Ward DS, Reznick RK, et al (2011) Faculty development in assessment: the missing link in competency-based medical education. *Acad Med* 86:460–7
- Jouquan J, Parent F (2015) Comment élaborer et analyser un référentiel de compétences en santé? De Boeck Supérieur, Bruxelles, pp 101–13
- Roegiers X (2000) Une pédagogie de l'intégration : compétences et intégration des acquis dans l'enseignement. De Boeck, Bruxelles, p66
- Scallon G (2007) L'évaluation des apprentissages dans une approche par compétences. De Boeck Supérieur, Bruxelles, pp 104–5
- Pastré P (2011) La didactique professionnelle. Presses Universitaires de France, Paris, 352p
- Ministère des Affaires sociales et de la Santé, ministère de la Défense, ministère de l'Enseignement supérieur et de la Recherche (2013) Arrêté du 8 avril 2013 relatif au régime des études en vue du premier et du deuxième cycle des études médicales. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027343762&categorieLien=id> (Dernier accès le 6 octobre 2019)
- Batalden P, Leach D, Swing S, et al (2002) General competencies and accreditation in graduate medical education. *Health Aff* 21:103–11
- Schuwirth LWT, van der Vleuten CPM (2004) Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 38:974–9
- Jouquan J (2002) L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie médicale* 3:38–52
- Brydges R, Hatala R, Zendejas B, et al (2015) Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med* 90:246–56
- Griswold S, Frallicardi A, Boulet J, et al (2018) Simulation-based education to ensure provider competency within the health care system. *Acad Emerg Med*. 25:168–76
- Stroben F, Schröder T, Dannenberg KA, et al (2016) A simulated night shift in the emergency room increases students' self-efficacy independent of role taking over during simulation. *BMC Med Educ* 16:177
- Ghazali A, Léger A, Petitpas F, et al (2016) Development and validation of a performance Assessment Scale for Chest Tube Insertion in traumatic pneumothorax. *J Pulm Respir Med* 6:346
- Maignan M, Koch FX, Chaix J, et al (2016) Team Emergency Assessment Measure (TEAM) for the assessment of non-technical skills during resuscitation: validation of the French version. *Resuscitation* 101:115–20
- Downing SM (2003) Validity: on meaningful interpretation of assessment data. *Med Educ* 37:830–7
- République française (2016) Décret n° 2016-1597 du 25 novembre 2016 relatif à l'organisation du troisième cycle des études de médecine et modifiant le Code de l'éducation. JORF n° 0276 du 27 novembre 2016. www.legifrance.gouv.fr/eli/decret/2016/11/25/MENS_16209996d:JO (Dernier accès le 19 février 2019).
- Riou B (2017) 2017: l'an 1 du diplôme d'études spécialisées de médecine d'urgence. *Ann Fr Med Urgence* 7:1–4
- Ministère des Affaires sociales et de la Santé (2018) Arrêté du 12 avril 2018 fixant la liste des recherches mentionnées au 3° de l'article L. 1121-1 du code de la santé publique https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFTEXT000036805820 (Dernier accès le 10 octobre 2018)
- Murphy MK, Black NA, Lamping DL, et al (1998) Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 2:1–88
- Jones J, Hunter D (1995) Consensus methods for medical and health services research. *BMJ* 311:376–80
- Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ (2017) Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med Teach* 39:14–9
- Collégiale des universitaires de médecine d'urgence (CNUMU), Collège national des enseignants de thérapeutiques (APNET) (2017) Urgences. Défaillances viscérales aiguës. Situations exceptionnelles. Med-Line. Paris, 493p

24. Allain M, Kuczer V, Longo C, et al (2018) Place de la simulation dans la formation initiale des urgentistes : enquête nationale observationnelle. *Ann Fr Med Urgence* 8:75–82
25. Granry JC (2012) Guide de bonnes pratiques en matière de simulation en santé. https://www.has-sante.fr/portail/jcms/c_1355826/fr/la-has-definit-les-bonnes-pratiques-pour-developper-des-programmes-de-simulation-en-sante (Dernière accès le 10 octobre 2018)
26. Motola I, Devine LA, Chung HS, et al (2013) Simulation in healthcare education: a best evidence practical guide. *AMEE Guide* n° 82. *Med Teach* 35:e1511–e30
27. Kolbe M, Grande B (2015) Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Pract Res Clin Anaesthesiol* 29:87–96
28. Wiggins G (1989) Teaching to the (authentic) test. *Educational Leadership* 46:41–7
29. Lemenu D, Heinen, E (2015) Comment passer des compétences à l'évaluation des acquis des étudiants. *De Boeck Supérieur, Bruxelles*, pp 50–6
30. Landeta J (2006) Current validity of the Delphi method in social sciences. *Technol Forecast Soc Change* 73:467–82