

## Évaluation objective structurée de l'animation du débriefing (ÉOSAD) : traduction, adaptation et validation d'une rubrique

### *Objective Structured Assessment of Debriefing: French translation, adaptation and validation of an assessment rubric*

Patrick LAVOIE<sup>1,2,3,\*</sup>, Marie-France DESCHÊNES<sup>1,2</sup>, Louise-Andrée BRIEN<sup>1,2</sup>, Mélanie RADERMAKER<sup>1,2</sup>, et Louise BOYER<sup>1,2</sup>

<sup>1</sup> Faculté des sciences infirmières, Université de Montréal, Montréal, Québec, Canada

<sup>2</sup> Centre d'innovation en formation infirmière, Université de Montréal, Montréal, Québec, Canada

<sup>3</sup> Centre de recherche, Institut de cardiologie de Montréal, Montréal, Québec, Canada

Manuscrit soumis à la rédaction le 17 juillet 2019; commentaires éditoriaux formulés aux auteurs le 5 février 2020 ; accepté pour publication le 11 février 2020

**Résumé – Contexte :** Le débriefing en simulation clinique requiert des habiletés d'animation de la part des formateurs. Bien que des outils d'évaluation validés existent et pourraient servir à leur formation, ils sont uniquement disponibles en anglais et ne sont pas adaptés à la réalité québécoise. **But :** (1) Traduire une rubrique d'évaluation de la qualité de l'animation du débriefing (*Objective Structured Assessment of Debriefing*) en français et l'adapter au contexte québécois ; et (2) en tester la validité et la fidélité dans un contexte de formation en sciences infirmières. **Méthodes :** La rubrique a été traduite et adaptée par trois traducteurs indépendants. Des experts en simulation ( $n=11$ ) en ont validé le contenu selon une méthode Delphi modifiée. La rubrique traduite a été mise à l'essai auprès d'animateurs en simulation ( $n=10$ ) qui ont évalué la qualité de débriefings filmés ( $n=16$ ). **Résultats :** La rubrique présente une forte validité de contenu selon les experts et une forte validité concomitante en comparaison avec un score d'évaluation générale du débriefing. Les résultats pour les fidélités interjuges et test-retest montrent une variabilité dans l'attribution des scores par les animateurs de débriefing. **Conclusion :** La rubrique fournit un langage et des critères communs pour évaluer la qualité d'animation des débriefings. Elle pourrait servir de base à la formation des formateurs en précisant des critères de performance et des descripteurs reflétant différents niveaux de performance. Les qualités psychométriques observées suggèrent toutefois qu'une formation plus exhaustive et des modifications dans les descripteurs seraient nécessaires avant que la rubrique ne soit utilisée en contexte de recherche.

**Mots clés :** débriefing, simulation, outil d'évaluation, validation, formation des formateurs

**Abstract. Context:** Educators must be skilled to facilitate debriefing after healthcare simulations. While tools to assess the quality of debriefing exist and could guide educators' training, they are only available in English and are not adapted to the reality of Quebec. **Purpose:** (1) To translate a rubric to assess the quality of debriefing (*Objective Structured Assessment of Debriefing*) in French and adapt it to the context of Quebec; and (2) to test the validity and fidelity of the translated and adapted rubric. **Methods:** Three independent translators translated and adapted the rubric. Simulation experts ( $n=11$ ) validated the rubric's content through a modified Delphi process. The translated rubric was piloted with a sample of simulation educators ( $n=10$ ) whom assessed the quality of recorded debriefings ( $n=16$ ). **Results:** The rubric presented strong content validity according to simulation experts and strong concurrent validity when compared to a general debriefing assessment score. Results for interrater and test-retest reliability showed some variability among educators. **Conclusion:** The rubric offers a common language and a set of criteria to assess the quality of debriefing. The rubric could guide educators' training by specifying performance criteria and descriptors reflecting various levels of performance. However, results regarding psychometric properties of the rubric suggest that extensive training and some modification to descriptors are required prior to its use as a research instrument.

**Keywords:** debriefing, simulation, assessment tool, validation, educator training

\*Correspondance et offprints: Patrick LAVOIE, Faculté des sciences infirmières, Pavillon Marguerite-d'Youville, CP 6128 succ. Centre-ville, H3C 3J7 Montréal, Québec, Canada.  
Mailto: [Patrick.lavoie.1@umontreal.ca](mailto:Patrick.lavoie.1@umontreal.ca).

**Tableau I.** Éléments d'évaluation de la qualité de l'animation du débriefing en simulation.

Outils	Dimensions
1. <i>Debriefing Assessment for Simulation in Healthcare</i> (DASH) [21]	Six éléments-clés d'un débriefing : 1. Établir un climat favorable à l'apprentissage ; 2. Maintenir un climat favorable à l'apprentissage ; 3. Conduire le débriefing de manière structurée ; 4. Susciter l'engagement dans l'échange ; 5. Identifier les écarts de performance et en analyser les raisons ; 6. Aider les apprenants à atteindre ou maintenir un bon niveau de performance future
2. <i>Objective Structured Assessment of Debriefing</i> (OSAD) [22]	Huit dimensions/éléments d'un débriefing efficace [31] : 1. Approche de l'animateur : établir et maintenir la relation. Utiliser une approche non menaçante et honnête pour créer un environnement psychologique sécuritaire ; 2. Climat propice à l'apprentissage : expliquer le but du débriefing. Clarifier les objectifs et les attentes des apprenants dès le début ; 3. Engagement des apprenants : encourager la participation des apprenants par des questions ouvertes. Inviter les apprenants à contribuer activement à la discussion ; 4. Réaction : explorer la réaction des apprenants à l'expérience. Gérer adéquatement les apprenants qui sont confus ou malheureux (« <i>confused and unhappy</i> ») ; 5. Description : encourager les apprenants à l'autoréflexion sur l'expérience au moyen d'une approche par étapes 6. Analyse : aider les apprenants à explorer les raisons et les conséquences de leurs actions. Faire des liens avec des expériences antérieures pour offrir des explications ; 7. Rétroaction : offrir une rétroaction objective sur les habiletés cliniques (techniques) et de travail d'équipe. Identifier les comportements positifs en plus des écarts de performance en ciblant les comportements modifiables ; 8. Application : renforcer les points-clés d'apprentissage identifiés par les apprenants. Souligner comment les stratégies d'amélioration peuvent être appliquées dans la pratique clinique future

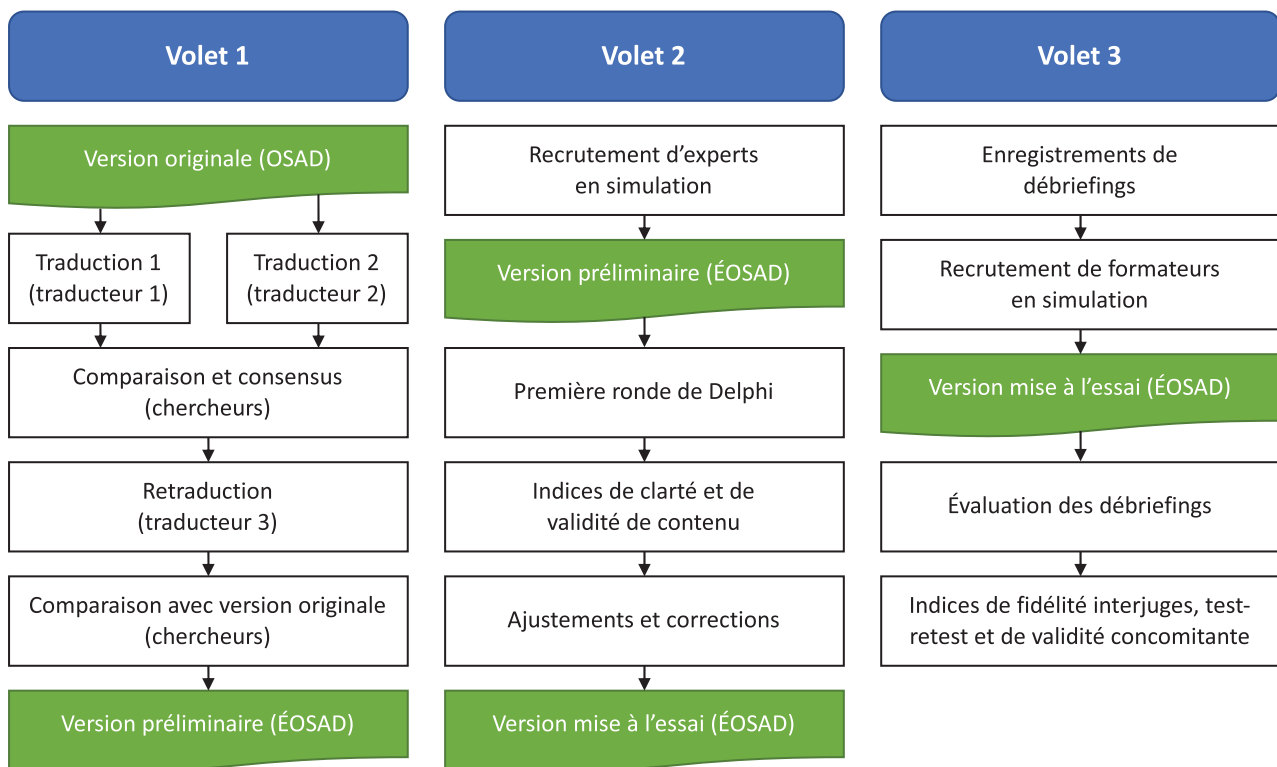
La traduction des éléments-clés du DASH est tirée de <https://harvardmedsim.org/debriefing-assessment-for-simulation-in-healthcare-dash-french/>. La traduction des dimensions/éléments de l'OSAD est celle des auteurs de la présente étude.

## Introduction

La simulation clinique est une stratégie pédagogique incontournable dans la formation initiale et continue des professionnels de la santé. La simulation est « *une technique, et non une technologie, pour remplacer ou amplifier des expériences réelles par des expériences guidées qui évoquent ou répliquent substantiellement des aspects de la réalité dans un format complètement interactif* » [1].

Le débriefing est une composante indissociable de la simulation clinique. Après une simulation, il est attendu que les apprenants disposeront de temps pour examiner, comprendre et apprendre de leur expérience [2,3]. Des études ont montré l'efficacité du débriefing pour améliorer la performance des apprenants [4] et ont établi que l'apprentissage en simulation peut être compromis en l'absence de débriefing [5,6]. Le débriefing a été défini de plusieurs manières : une discussion guidée par un facilitateur ou des pairs [7] ; un processus où des apprenants et un formateur réexaminent une expérience clinique [8] ou une réflexion guidée ou facilitée dans le cycle de l'apprentissage expérientiel [9]. Toutes ces définitions ont au moins un point en commun : elles impliquent une forme d'accompagnement ou de facilitation, c'est-à-dire un processus par lequel un animateur guide les apprenants dans l'analyse de leur expérience en simulation [10].

À cet égard, des standards de pratique pour la formation en simulation mentionnent que les animateurs de débriefing devraient être formés et compétents [2,11,12]. Les quelques données scientifiques existantes font référence à des habiletés pour structurer le débriefing et pour promouvoir la participation active des apprenants [13–18]. Or, le manque d'habiletés des animateurs est fréquemment soulevé comme un obstacle à l'intégration de la simulation dans les programmes de formation [19]. Par ailleurs, peu d'écrits en sciences de la santé offrent un soutien théorique pour évaluer et guider la formation des animateurs de débriefing [20]. Néanmoins, deux outils d'évaluation de la qualité de l'animation du débriefing sont fréquemment cités dans les écrits : le *Debriefing Assessment for Simulation in Healthcare* (DASH) [21] et l'*Objective Structured Assessment of Debriefing* (OSAD) [22]. Chacun propose un ensemble d'éléments sur lesquels on peut appuyer l'évaluation de la qualité du débriefing (Tab. I). Ces outils ont des qualités psychométriques comparables [21,22]. L'OSAD a obtenu un indice de validité de contenu de 0,94 et le DASH a été révisé par un panel d'experts en simulation. L'OSAD a obtenu un indice de fidélité inter-juges supérieur à celui du DASH (coefficients de corrélation intra-classe [ICC] = 0,88 et 0,74, respectivement). La fidélité test-retest de l'OSAD a également été montrée (ICC = 0,90). Pour ce qui est des indices de validité de construit, les scores à l'OSAD sont



OSAD : *Objective Structured Assessment of Debriefing*

ÉOSAD : Évaluation Objective Structurée de l'Animation du Débriefing

**Figure 1.** Principales étapes des trois volets de l'étude. OSAD : *Objective Structured Assessment of Debriefing*; ÉOSAD : Évaluation objective structurée de l'animation du débriefing.

fortement corrélés avec la perception générale de l'efficacité des débriefings ( $r=0,68$ ;  $p<0,01$ ), tandis que les scores au DASH varient significativement selon la qualité du débriefing telle qu'elle est perçue par des experts ( $F(2,226) = 486,2$ ;  $p<0,001$ ).

Néanmoins, le DASH apparaît limité dans sa capacité à offrir une rétroaction qui soutient l'apprentissage des habiletés d'animation de débriefing. Selon Feeney [23], une rétroaction à l'intention des formateurs devrait être basée sur des données observables, identifier les caractéristiques d'un accompagnement efficace et promouvoir la réflexion afin de soutenir l'amélioration de l'enseignement. En ce sens, il est suggéré d'employer une rubrique qui définit explicitement des descripteurs pour différents niveaux de performance. Cela a l'avantage d'indiquer clairement les différences entre les niveaux et d'offrir un langage commun aux formateurs [24]. Comme le DASH se fonde sur une évaluation quantitative des dimensions énumérées dans le tableau I (de 1 — extrêmement inefficace à 7 — extrêmement efficace), les différences entre les scores relèvent de la subjectivité de l'évaluateur et sont moins fondées sur des critères explicites.

L'OSAD pallie cette difficulté et se démarque par son potentiel à offrir aux animateurs une rétroaction qui pourrait contribuer au développement de leurs habiletés en matière de débriefing. Il comprend un manuel

d'utilisation et définit précisément huit critères permettant d'évaluer la qualité du débriefing. Chacun des critères (Tab. I) est évalué à l'aide d'une échelle de type Likert à cinq points allant de 1 (qualité minimale) à 5 (qualité maximale). Des descripteurs pour les scores 1, 3 et 5 sont fournis pour guider l'attribution des scores; les scores 2 et 4 ne comprennent pas de descripteurs et sont utilisés si la performance de l'animateur se situe entre deux descripteurs. Le score total peut donc varier de 8 à 40; un score élevé indique que l'animation du débriefing est de meilleure qualité.

Or, l'OSAD est uniquement disponible en anglais et a été développé au Royaume-Uni, un contexte différent de celui qui prévaut au Québec. En vue d'offrir un outil d'évaluation de l'animation du débriefing en français et cohérent avec le contexte de formation québécois, les objectifs de cette étude étaient de :

- traduire l'OSAD en français et de l'adapter au contexte québécois;
- tester la validité et la fidélité de la version traduite et adaptée.

Le but de cet article est d'explicitier les différentes étapes menées par l'équipe de recherche et ce, afin que l'OSAD puisse être utilisé à des fins de développement des habiletés d'animation du débriefing en contexte de simulation en sciences de la santé au Québec.

## Méthodes

En prenant appui sur les recommandations de Sousa et Rojjanasrirat [25], l'étude s'est déclinée en trois volets (Fig. 1) :

- traduction et adaptation ;
- validation de contenu ;
- mise à l'essai auprès de la population cible.

Ces recommandations ont été retenues pour guider l'étude puisqu'elles ont été formulées à partir d'une synthèse des principales approches méthodologiques pour la traduction, l'adaptation et la validation transculturelle d'instruments de mesure pour la recherche en sciences de la santé.

### Volet 1–Traduction et adaptation

Le premier volet de l'étude s'est appuyé sur les quatre premières étapes de la méthode recommandée par Sousa et Rojjanasrirat [25]. Deux traducteurs bilingues indépendants ont traduit le manuel d'utilisation (instructions, définitions des critères, exemples) et les descripteurs de la rubrique de l'anglais vers le français (étape 1). Les deux traducteurs étaient principalement francophones ; le traducteur 1 était familier avec le domaine du débriefing en simulation, alors que le traducteur 2 était familier avec le contexte de la formation au Québec, sans être familier avec le débriefing en simulation. Les deux versions françaises de l'OSAD ont ensuite été comparées par l'équipe de recherche pour déterminer les ambiguïtés et les différences dans les mots, les phrases et le sens employés (étape 2). À la suite d'un consensus au sein de l'équipe de recherche, une troisième version française de l'OSAD a été formulée et retraduite vers l'anglais par un troisième traducteur principalement anglophone et familier avec le contexte de la formation au Québec (étape 3). La version retraduite a été comparée avec la version anglaise originale de l'OSAD par l'équipe de recherche (étape 4). Cette comparaison a permis d'établir les équivalences conceptuelles, sémantiques et de contenu de la version traduite et adaptée et de la rubrique originale afin de consolider une version préliminaire de la rubrique ÉOSAD en français (Évaluation objective structurée de l'animation du débriefing).

### Volet 2–Validation de contenu

La méthode Delphi modifiée [26] a été utilisée pour réaliser la cinquième étape de la méthode de Sousa et Rojjanasrirat [25]. L'objectif de ce volet de l'étude était d'obtenir un consensus sur la clarté et la pertinence du manuel d'utilisation et des descripteurs de la version préliminaire de l'ÉOSAD. Une stratégie de recrutement par boule de neige a été employée. Les critères d'inclusion étaient :

- être un chercheur ou un formateur québécois qui s'intéresse à la simulation clinique ;

- utiliser la simulation depuis au moins cinq ans auprès d'étudiants ou de professionnels de la santé, critère précédemment utilisé pour définir l'expertise en simulation [21] ;
- comprendre et écrire le français.

L'équipe de recherche a contacté des individus représentant différentes disciplines des sciences de la santé qui emploient la simulation et le débriefing. S'ils souhaitaient participer à l'étude, les participants potentiels étaient invités à contacter l'équipe de recherche afin d'obtenir un lien pour accéder à un questionnaire en ligne.

Le questionnaire comprenait des questions sur les données sociodémographiques des experts, sur le manuel d'utilisation et sur les descripteurs de la rubrique. Les questions sur les données sociodémographiques ciblaient l'âge, la discipline, l'activité principale, le nombre d'années d'expérience en enseignement et avec la simulation, et les connaissances sur la simulation et le débriefing (évaluées sur une échelle de 0–aucune connaissance à 10–énormément de connaissances). Pour le manuel d'utilisation, les experts étaient invités à se prononcer sur la clarté (clair/pas clair) des instructions, des définitions et des exemples pour chaque critère. Pour les 24 descripteurs de la rubrique, les experts devaient se prononcer sur leur clarté, mais aussi sur leur pertinence (1–pas pertinent ; 2–peu pertinent ; 3–pertinent ; 4–très pertinent). Les experts pouvaient suggérer l'ajout ou le retrait d'éléments à la rubrique et inscrire des commentaires au besoin.

Les données sociodémographiques ont été analysées par des statistiques descriptives. Ensuite, des indices de clarté (IC ; pourcentage d'experts jugeant qu'un élément était clair) ont été calculés pour chaque élément (manuel et rubrique). Des indices de validité de contenu (IVC ; pourcentage d'experts jugeant un descripteur pertinent ou très pertinent) ont été calculés pour les descripteurs de la rubrique. L'IVC global de la rubrique (IVC-S ; moyenne de tous les IVC) a également été calculé. Des IC, IVC et IVC-S supérieurs à 0,78 étaient visés, ce qui correspond à au moins 78 % des experts jugeant qu'un élément est clair ou pertinent [27]. Des rondes supplémentaires du processus Delphi étaient prévues jusqu'à ce que les IC et IVC de chacun des éléments du manuel et de la rubrique atteignent le seuil recommandé et que les experts ne suggèrent plus de changements majeurs. Après chaque ronde du Delphi, l'équipe de recherche a apporté des corrections à la version préliminaire de l'ÉOSAD pour arriver à la version mise à l'essai. Les analyses statistiques ont été réalisées avec le logiciel SPSS version 24.

### Volet 3–Mise à l'essai auprès de la population cible

La sixième étape de la méthode de Sousa et Rojjanasrirat [25] a été adaptée pour le troisième volet de l'étude, qui visait à tester la validité et la fidélité de l'ÉOSAD. Pour ce faire, 16 vidéos de débriefings ont été filmées à la suite de simulations dans le cadre de deux cours au baccalauréat en sciences infirmières (soins critiques et

périnatalité). Les animateurs de ces débriefings ont été informés de l'étude par courriel, en spécifiant qu'ils étaient libres de participer ou non. S'ils acceptaient que leur débriefing soit enregistré, l'étude était présentée aux étudiants qui s'apprêtaient à participer aux simulations, qui étaient eux aussi libres d'accepter de participer ou non. Dans l'affirmative, le débriefing était enregistré à l'aide d'une caméra installée dans la salle de débriefing.

Par la suite, d'autres animateurs de débriefing ont été recrutés pour utiliser l'ÉOSAD afin d'évaluer la qualité des débriefings préalablement enregistrés. Le recrutement a été réalisé au moyen d'une annonce publiée sur le site internet d'une faculté des sciences infirmières et d'un courriel aux étudiants aux cycles supérieurs de cette faculté. Un calcul de taille d'échantillon [28] réalisé à partir des résultats de fidélité inter-juges de la version originale de la rubrique (ICC 0,88) a déterminé que 10 animateurs évaluant 16 débriefings permettraient d'identifier un ICC minimal de 0,75 (intervalle de confiance à 95 % [IC 95 %] : 0,60–0,88). Les animateurs ont été rencontrés pour examiner le manuel d'utilisation, l'ÉOSAD et les questionnaires de l'étude, et poser des questions à l'équipe de recherche. Ils ont signé un formulaire de consentement et un engagement de confidentialité quant aux contenus des vidéos. Ils ont été avisés de réaliser l'évaluation de manière individuelle et de ne communiquer entre eux à ce sujet sous aucun prétexte.

Les données pour le volet 3 de l'étude ont entièrement été collectées de manière électronique. Dans un premier temps de mesure, chaque animateur a rempli un questionnaire sociodémographique et 16 questionnaires sur les vidéos des débriefings. Le questionnaire sociodémographique était identique à celui des experts (volet 2). Les questionnaires sur les vidéos de débriefing comprenaient neuf questions : les huit premières présentaient les critères de l'ÉOSAD et les descripteurs associés, la neuvième touchait l'évaluation générale de la qualité de l'animation du débriefing (échelle allant de 1 — pas bonne du tout à 7 — excellente). Le second temps de mesure a eu lieu un mois après le premier. À ce moment, tous les animateurs ont rempli à nouveau un questionnaire pour une des vidéos de débriefing qu'ils avaient évaluées lors du premier temps de mesure. La vidéo utilisée lors du deuxième temps de mesure a été sélectionnée au hasard et était la même pour tous les animateurs.

Les données sociodémographiques ont été analysées à l'aide de statistiques descriptives. Les données du premier temps de mesure ont servi à tester la fidélité inter-juges et la validité concomitante du score total sur la rubrique. La fidélité inter-juges a d'abord été calculée à l'aide d'un ICC basé sur un modèle mixte à deux facteurs mesurant l'accord absolu entre les experts. Puis, selon les recommandations de Stemler [29] pour l'évaluation des propriétés psychométriques de rubriques, deux aspects de la fidélité inter-juges ont été testés, soit le consensus (accord exact entre les juges) et la cohérence (constance dans l'application des critères, sans exigence de consensus; un juge pourrait systématiquement attribuer des scores plus bas comparativement à un autre juge). Le consensus a été

mesuré par les pourcentages d'accord exact et d'accord adjacent pour chaque critère. Pour les pourcentages d'accord adjacent, le pourcentage de scores supérieurs ou inférieurs d'un point à la médiane pour chaque critère a d'abord été calculé pour chaque vidéo, pour ensuite calculer la moyenne des pourcentages sur un même critère pour toutes les vidéos. Des pourcentages d'accord exact supérieurs à 70 % et adjacent supérieurs à 90 % ont été considérés adéquats [29]. La cohérence a été mesurée au moyen du coefficient alpha de Cronbach ( $\alpha$ ); des valeurs supérieures à 0,70 étaient visées.

Les données du second temps de mesure ont été comparées aux données du premier temps pour tester la fidélité test-retest de la rubrique. Les mêmes indicateurs ont été employés (ICC,  $\alpha$ , pourcentages d'accord exact et d'accord adjacent). La cohérence individuelle de chaque animateur a également été mesurée ( $\alpha$ ). Quant à la validité concomitante, elle a été évaluée à partir de la corrélation de Spearman entre le score total sur l'ÉOSAD et le score d'évaluation générale du débriefing; un seuil de signification bilatéral de 0,05 a été choisi.

## Résultats

### Volet 1—Traduction et adaptation

Lors de la traduction et de l'adaptation de la rubrique originale, les mots « *facilitator* » et « *learners* » ont été traduits par « animateur » et « apprenants ». Les répétitions du mot « *facilitator* » dans les descripteurs ont été retirées pour alléger le texte. Une uniformisation du vocabulaire et de la structure des descripteurs a été effectuée afin d'assurer la constance dans la formulation des descripteurs et pour en clarifier les distinctions. Certains concepts ont occasionné des difficultés particulières lors de la traduction. Pour le concept anglophone « *rapport* » (critère 1), les traducteurs et l'équipe de recherche ont hésité entre les termes « relation », « interaction » et « bons rapports », pour finalement retenir ce dernier. Le critère 7 « *diagnosis* » a été traduit et adapté par « rétroaction ». Pour les concepts « *clinical and teamwork skills* » (critère 7), l'équipe de recherche a retenu « habiletés cliniques » et « habiletés de collaboration ».

### Volet 2—Validation de contenu

Sur 28 individus contactés, 11 experts en simulation ont accepté de participer à la validation de contenu de l'ÉOSAD (39 %). Les experts provenaient de disciplines variées : inhalothérapie, médecine, physiothérapie et sciences infirmières. En moyenne, ils comptaient huit ans d'expérience en simulation ( $\pm 3$  ans) et évaluaient leurs connaissances en matière de simulation et de débriefing à 8/10. Leurs caractéristiques sociodémographiques sont présentées dans le [tableau II](#).

Le manuel d'utilisation comprenait des instructions, les définitions des critères et des exemples de situations associées à ceux-ci. Les IC des instructions et des définitions variaient entre 0,91 et 1,00, ce qui indique que 91 à 100 % des experts jugeaient qu'elles étaient



**Tableau II.** Caractéristiques sociodémographiques des experts et des animateurs.

	Experts ( <i>n</i> = 11)	Formateurs ( <i>n</i> = 10)
Âge (années)	44,5 (9,9)	29,0 (4,7)
Discipline		
Inhalothérapie	1	0
Médecine	2	0
Physiothérapie	2	0
Sciences infirmières	6	10
Activité principale		
Enseignement	5	1
Gestion	1	1
Recherche	3	4
Soins aux patients	2	4
Expérience d'enseignement (années)	16,1 (7,9)	2,7 (3,1)
Expérience simulation (années)	8,1 (3,2)	2,4 (1,9)
Connaissances simulation (/10)	8,0 (0,8)	7,4 (1,0)
Connaissances débriefing (/10)	7,6 (1,3)	6,7 (1,2)

Les données sont des moyennes (écarts-types) ou des fréquences.

claires. Toutefois, bien que les IC des exemples variaient entre 0,82 et 1,00, les experts ont commenté qu'ils étaient source de confusion et sujets à critiques (ex. : risque de biais, caricatures, exagération). D'un commun accord, l'équipe de recherche a décidé de retirer les exemples du manuel et de conserver uniquement les instructions et les définitions des critères.

La majorité des IC des descripteurs de la rubrique ont atteint 0,82–1,00, à l'exception de deux descripteurs ayant obtenu des IC de 0,73. Il s'agissait de descripteurs pour les critères 2–climat propice à l'apprentissage et 7–rétroaction. Pour le critère 2, les experts estimaient que la distinction entre le « but du débriefing », les « attentes des apprenants » et les « objectifs d'apprentissage » n'était pas claire. Des modifications mineures ont été apportées afin de clarifier ces distinctions et de maximiser les différences entre les descripteurs. Pour ce qui est du critère 7, les experts ont critiqué l'inclusion de seulement deux types d'habiletés (techniques et de collaboration). Le terme « habiletés de collaboration » a été remplacé par « points variés », afin d'être plus inclusif. En effet, la rétroaction en débriefing peut être offerte sur divers éléments, par exemple les habiletés cliniques, de communication, d'enseignement auprès du patient et d'intervention [30].

Les IVC des descripteurs variaient entre 0,91 et 1,00, dépassant le seuil visé de 0,78. L'IVC-S de la rubrique entière a atteint 0,99. Considérant ces indices de validité de contenu élevés, une seule ronde de Delphi a été réalisée. La version mise à l'essai de l'ÉOSAD est présentée dans le [tableau III](#).

### Volet 3–Mise à l'essai auprès de la population cible

Dix animateurs de débriefing ont participé au dernier volet de l'étude. En moyenne, ils comptaient moins de trois

ans d'expérience en enseignement et en simulation. Malgré cela, ils estimaient que leurs connaissances en simulation et en débriefing se situaient, en moyenne, entre 6 et 8/10. Leurs caractéristiques sociodémographiques sont présentées dans le [tableau II](#). La plupart des animateurs œuvraient en pratique clinique ou étaient en cours d'études. En effet, les quatre animateurs qui ont indiqué que leur activité principale était la recherche réalisaient des études scientifiques dans le cadre de leur programme de maîtrise.

Pour la rubrique complète, l'ICC de fidélité inter-juges a atteint 0,34 (IC 95 % : 0,18–0,59), ce qui est inférieur à l'ICC de la rubrique originale (0,88). Au niveau des critères, les résultats pour la fidélité inter-juges sont présentés dans le [tableau IV](#). Du point de vue de la cohérence, cinq critères (1, 2, 4, 5 et 8) ont obtenu un coefficient  $\alpha$  supérieur au seuil de 0,70 et deux critères s'en approchent (3 et 6). Un seul critère (7) a obtenu un coefficient  $\alpha$  plus faible de 0,50. Les animateurs ont ainsi appliqué la majorité des critères avec constance, à l'exception du critère 7 qui a été appliqué de manière plus variable. Du point de vue du consensus, aucun critère n'a atteint le seuil d'accord exact souhaité (70 %), ce qui peut expliquer le faible ICC obtenu. Par contre, tous les critères ont atteint (ou approchent) le seuil d'accord adjacent de 90 %. Ceci indique que pour un même débriefing, les animateurs octroyaient rarement le même score pour un critère donné, mais que les scores se situaient majoritairement dans un intervalle de plus ou moins un point.

Les résultats pour la fidélité test-retest de l'ÉOSAD sont présentés dans le [tableau V](#). L'ICC test-retest pour l'ensemble de la rubrique a atteint 0,46 (IC 95 % : 0,04–0,74), ce qui est également inférieur à l'ICC test-retest obtenu par la rubrique originale (0,90). Sur le plan individuel, une majorité d'animateurs (*n* = 7) a montré une cohérence acceptable dans l'application des critères ( $\alpha$  entre 0,73 et 0,93). Trois animateurs ont toutefois présenté une cohérence plus faible ( $\alpha$  de 0,60, 0,33 et  $-0,18$ ), indiquant qu'ils n'ont pas appliqué les critères de la même façon lors de la première et de la seconde évaluation des débriefings. Les données de ces trois animateurs ont donc contribué à diminuer les indices de fidélité test-retest. Le [tableau V](#) présente à la fois les résultats de fidélité test-retest pour tous les animateurs (*n* = 10) et pour les animateurs ayant montré une cohérence acceptable ( $\alpha > 0,70$ ; *n* = 7). La plupart des critères (1 à 5) montrent un degré de cohérence acceptable. Les critères 6, 7 et 8 ont toutefois été évalués de manière plus variable de la première à la seconde observation. Comme c'était le cas pour la fidélité inter-juges, une grande majorité de critères n'a pas atteint le seuil d'accord exact souhaité. Bien que les pourcentages d'accord adjacent soient plus élevés, certains n'ont pas atteint le seuil de 90 %. Toutefois, ces résultats représentent un nombre plus restreint d'observations (*n* = 20 pour tous les animateurs; *n* = 14 pour les animateurs avec cohérence acceptable).

En ce qui a trait à la validité concomitante, les scores totaux à l'ÉOSAD ont atteint 29,94 ( $\pm 4,94$ ) et les scores d'évaluation globale des débriefings ont atteint 5,29 ( $\pm 1,13$ ) en moyenne. Une forte corrélation entre ces deux variables a été observée ( $r_s = 0,74$ ;  $p < 0,001$ ).

**Tableau III.** Rubrique d'Évaluation objective structurée de l'animation du débriefing (ÉOSAD).

Critères	1	2	3	4	5
1. Approche de l'animateur	Adopte une approche moralisatrice, axée sur la confrontation		Tente d'établir de bons rapports avec les apprenants, mais son approche est trop critique ou trop familière		Établit et maintient de bons rapports avec les apprenants ; son approche est honnête sans être menaçante, ce qui crée un environnement psychologique sécuritaire
2. Climat propice à l'apprentissage	Ne clarifie pas les attentes des apprenants, ni leur implication attendue ou le déroulement du débriefing		Explique le but du débriefing, mais ne clarifie pas quelles sont les attentes des apprenants		Explique le but du débriefing et clarifie quels sont les attentes et les objectifs d'apprentissage des apprenants dès le départ
3. Engagement des apprenants	Monopolise la parole, ne cherche pas à impliquer tous les apprenants ; le débriefing prend l'allure d'un cours magistral		Implique les apprenants, mais surtout par des questions fermées ; ne sollicite pas les apprenants plus passifs		Favorise la participation de tous les apprenants par des questions ouvertes ; encourage leur contribution active
4. Réaction	N'explore pas les réactions des apprenants ni leurs émotions à la suite de la simulation		Demande aux apprenants comment ils se sentent, mais explore peu leurs réactions		Explore en profondeur les réactions des apprenants et gère adéquatement les émotions négatives ou excessives
5. Description	N'invite pas les apprenants à décrire ce qui s'est passé pendant la simulation		Invite les apprenants à décrire ce qui s'est passé pendant la simulation, mais intervient fréquemment pour donner son avis		Invite les apprenants à décrire ce qui s'est passé pendant la simulation, étape par étape, selon leurs perspectives
6. Analyse	N'explore pas les raisons ni les conséquences des actions des apprenants		Interprète lui-même les actions des apprenants et leurs conséquences ; peu d'occasion pour les apprenants de créer des liens avec des expériences antérieures		Aide les apprenants à explorer les raisons qui justifient leurs actions et leurs conséquences, en ciblant des exemples spécifiques, et en encourageant la création de liens avec des expériences antérieures
7. Rétroaction	N'offre pas de rétroaction aux apprenants ; ne les aide pas à identifier leurs forces ni leurs points à améliorer		Offre une rétroaction centrée sur les habiletés cliniques (techniques) ; discute des erreurs et non des comportements modifiables		Offre une rétroaction objective sur des points variés ; identifie autant des forces que des points à améliorer ; cible des comportements modifiables
8. Application	N'offre pas d'occasion aux apprenants d'identifier leurs apprentissages ou des stratégies d'amélioration à appliquer dans le futur		Discute brièvement des apprentissages réalisés et des stratégies d'amélioration avec les apprenants, mais aborde peu les moyens de les appliquer dans le futur		Renforce les apprentissages clés identifiés par les apprenants et les encourage à s'exprimer sur les moyens de les appliquer dans le futur

## Discussion

Cette étude a permis de traduire en français et d'adapter au contexte québécois une rubrique d'évaluation de la qualité de l'animation du débriefing en simulation auprès de professionnels de la santé. Les résultats montrent que la rubrique présente une forte validité de contenu selon des experts de disciplines variées et une forte validité concomitante, en comparaison avec un score d'évaluation générale du débriefing. En fait, les indices de validité de l'ÉOSAD ont dépassé ceux de la rubrique originale (IVC-S = 0,94 et  $r = 0,68$ ,  $p < 0,01$ ). Ceci indique que l'ÉOSAD mesure bel et bien la qualité de l'animation du débriefing, ce pour quoi il a été conçu.

Toutefois, les résultats pour les fidélités inter-juges et test-retest de l'ÉOSAD sont plus mitigés. De manière générale, les analyses révèlent que l'accord exact entre les animateurs était rare. En effet, les scores tendaient à varier de plus ou moins un point pour chacun des critères observés ; il en est de même pour les scores attribués par les animateurs lorsqu'ils évaluaient un même débriefing à deux occasions séparées par un intervalle d'un mois. Ce phénomène est potentiellement attribuable à plusieurs facteurs. D'abord, la rubrique comprend des descripteurs pour seulement trois des cinq niveaux, laissant ainsi deux niveaux intermédiaires sans descripteurs. Ceci pourrait avoir engendré une variation dans les scores octroyés, qui s'observe notamment par le faible pourcentage d'accord

**Tableau IV.** Fidélité inter-juges de l'évaluation objective structurée de l'animation du débriefing (ÉOSAD) (version française).

	Médiane (étendue)	$\alpha$	Accord exact <sup>a</sup>	Accord adjacent <sup>a</sup>
Total	30 (15–40)	0.89	–	–
Critère 1	5 (2–5)	0.74	<b>61</b>	92
Critère 2	2 (1–5)	0.95	<b>63</b>	<b>85</b>
Critère 3	4 (3–5)	<b>0.68</b>	<b>59</b>	94
Critère 4	3 (1–5)	0.84	<b>53</b>	91
Critère 5	4 (2–5)	0.72	<b>54</b>	91
Critère 6	4 (1–5)	<b>0.67</b>	<b>53</b>	91
Critère 7	4 (2–5)	<b>0.50</b>	<b>52</b>	93
Critère 8	4 (1–5)	0.74	<b>48</b>	<b>85</b>

<sup>a</sup> Les données sont des pourcentages. Les données en gras n'ont pas atteint les seuils souhaités.

**Tableau V.** Fidélité test-retest de l'évaluation objective structurée de l'animation du débriefing (ÉOSAD) (version française).

	$\alpha$	Accord exact <sup>a</sup>	Accord adjacent <sup>a</sup>
Tous les critères	0.74 (0.99)	<b>58 (63)</b>	<b>87 (91)</b>
Critère 1	<b>0.68 (0.81)</b>	70 (79)	90 (93)
Critère 2	0.76 (0.83)	<b>40 (43)</b>	<b>65 (79)</b>
Critère 3	<b>0.68 (0.87)</b>	<b>60 (71)</b>	95 (100)
Critère 4	0.76 (0.77)	<b>65 (57)</b>	90 ( <b>86</b> )
Critère 5	0.80 (0.80)	<b>55 (50)</b>	100 (100)
Critère 6	<b>0.62 (0.67)</b>	<b>60 (71)</b>	95 (93)
Critère 7	<b>0.02 (0.66)</b>	<b>65 (79)</b>	<b>85 (93)</b>
Critère 8	<b>0.12 (0.50)</b>	<b>45 (50)</b>	<b>75 (86)</b>

<sup>a</sup> Les données sont des pourcentages. Les données en gras n'ont pas atteint les seuils souhaités. Les valeurs entre parenthèses représentent les résultats lors du retrait des trois animateurs ayant montré une cohérence plus faible.

exact et le pourcentage élevé d'accord adjacent. Cette observation corrobore certaines données antérieures par rapport à l'utilisation de rubriques qui stipulent que les pourcentages d'accord exact atteignent rarement le seuil souhaité de 70 %, mais que les pourcentages d'accord adjacent sont généralement plus satisfaisants (> 90 %) [24]. Ainsi, une réduction du nombre de niveaux d'une rubrique mènerait généralement à une amélioration des fidélités inter-juges et test-retest [24].

Un autre facteur pourrait être la sélection et la formation des animateurs ayant utilisé la rubrique (volet 3). Dans cette étude, les animateurs ont participé à une rencontre qui visait principalement à valider leur compréhension des critères et des descripteurs. Aucun exercice n'a été réalisé pour qu'ils s'entraînent à utiliser la rubrique et qu'ils reçoivent une rétroaction. Par ailleurs, ces animateurs étaient représentatifs de la population qui est susceptible d'utiliser la rubrique, soit des formateurs avec une expérience restreinte en débriefing. Dans l'étude originale de validation de l'OSAD [22], la formation des évaluateurs n'a pas été décrite, outre le fait qu'ils avaient une expérience en éducation. Ainsi, il est raisonnable de penser que leur expérience et leur formation aient influencé la manière dont ils ont complété la rubrique. Une formation plus complète, incluant des exercices pratiques, serait susceptible de réduire les écarts entre les scores

octroyés [24]. Il serait aussi intéressant de réaliser une autre étude afin de vérifier si l'expérience en débriefing des utilisateurs de l'ÉOSAD influence les résultats relatifs à la fidélité de la rubrique.

Par ailleurs, les critères 2, 7 et 8 ont semblé présenter des difficultés en termes de fidélité. Des problèmes de clarté avaient d'ailleurs été soulevés par les experts lors du volet 2 quant à leur formulation. Comparativement aux autres critères, les descripteurs incluaient plusieurs points d'évaluation, ce qui pourrait expliquer les variations observées. Par exemple, pour le critère 8 (application), il est question de l'identification des apprentissages, des stratégies d'amélioration et des moyens de les appliquer dans le futur. Une avenue à explorer serait de réduire le nombre de points à évaluer par critère, mais ceci serait susceptible de réduire la richesse de la rétroaction offerte aux animateurs évalués. Il serait également possible d'améliorer la définition des critères afin de clarifier le sens de chacun des points à évaluer. Par ailleurs, la sous-division de ces critères pourrait être envisagée, mais l'accroissement du nombre de critères pourrait augmenter la complexité de la rubrique et son temps de complétion. De plus, pour le critère 2 (climat propice à l'apprentissage), il est possible que la distinction entre les attentes des apprenants, leurs objectifs d'apprentissage et les buts du débriefing n'ait pas été suffisamment explicite malgré les



modifications proposées par l'équipe de recherche. Pour le critère 7 (rétroaction), le fait d'utiliser l'expression « points variés » plutôt que de cibler des points de rétroaction précis — comme c'était le cas dans la version originale de l'OSAD — a pu contribuer à diminuer la précision des descripteurs. Une formation auprès des utilisateurs de la rubrique pourrait être une avenue pour clarifier les distinctions entre les descripteurs et favoriser les fidélités inter-juges et test-retest.

Cette étude comporte certaines limites. D'abord, les modifications proposées par l'équipe de recherche lors de la traduction de la rubrique n'ont pas été soumises aux différents traducteurs ayant été impliqués, ce qui aurait pu permettre de les valider. Ensuite, un seul traducteur a retraduit la rubrique proposée du français vers l'anglais, contrairement aux recommandations de Sousa et Rojjanasrirat [25] qui suggèrent d'employer deux traducteurs. Bien que des changements mineurs aient été effectués dans la rubrique à la suite de la première ronde de Delphi, l'équipe de recherche a choisi de ne pas procéder à une seconde ronde en raison des indices de validité de contenu élevés obtenus. Pour ce qui est de la mise à l'essai, un échantillon avec des caractéristiques particulières a été formé dans un seul milieu d'enseignement. Cet échantillon pourrait ne pas être représentatif d'autres contextes. En outre, des actions pourraient être mises en place pour améliorer la rubrique, notamment la réduction du nombre de niveaux d'évaluation, la division des critères pour réduire le nombre de points d'évaluation et la formation des utilisateurs. Toutes ces modifications impliqueraient néanmoins des impératifs pour reprendre le processus de validation (consultation d'experts et mise à l'essai de la rubrique modifiée). De plus, il serait pertinent de poursuivre la validation de l'ÉOSAD dans des conditions différentes de celles décrites dans la présente étude, notamment à la suite d'une formation plus exhaustive incluant des exercices d'utilisation de la rubrique suivis de rétroaction auprès des utilisateurs.

## Conclusion

Il s'agit à notre connaissance du premier outil d'évaluation de la qualité du débriefing ayant été traduit en français, adapté au contexte québécois et mis à l'essai afin d'en tester la validité et la fidélité dans un contexte de formation en sciences infirmières. L'implication d'experts de plusieurs disciplines des sciences de la santé vient renforcer son potentiel d'utilisation. Considérant les résultats obtenus, il semble que la rubrique ne puisse pas encore être utilisée dans un contexte de recherche ou d'évaluation à enjeux élevés. En effet, les résultats montrent que les scores attribués sont sujets à une certaine variation selon les évaluateurs et le moment d'évaluation. Toutefois, puisque cette variation est demeurée somme toute acceptable, l'outil se présente comme une avenue prometteuse pour l'accompagnement des animateurs dans une logique d'amélioration continue de la qualité des débriefings. En effet, l'ÉOSAD fournit un

langage et des critères communs pour évaluer la qualité d'animation des débriefings dans un contexte uni- ou interdisciplinaire. En outre, l'ÉOSAD pourrait également servir de base à la formation des animateurs en précisant des critères de performance et des descripteurs reflétant différents niveaux de performance.

## Contributions

Patrick Lavoie et Marie-France Deschênes ont conçu le protocole de recherche. Tous les auteurs ont participé au recueil des données et à l'interprétation des résultats. L'analyse statistique a été effectuée par Patrick Lavoie, qui a rédigé le manuscrit avec la collaboration de Mélanie Radermaker. Tous les auteurs ont révisé le manuscrit, contribué à son contenu et approuvé la version soumise.

## Conflit d'intérêts

Les auteurs n'ont pas de conflit d'intérêts à déclarer.

## Approbation éthique

L'étude a été approuvée par le Comité d'éthique de la recherche en santé de l'Université de Montréal en date du 23 janvier 2019 (CERSES-18-012-D).

## Financement

Ces travaux ont été rendus possibles grâce au soutien de l'Équipe FUTUR, une infrastructure de recherche subventionnée par le Fonds de recherche du Québec – Société et culture (2017–2021).

## Remerciements

Les auteurs souhaitent remercier Sonal Arora, auteure originale de l'OSAD, pour l'autorisation de traduire et d'adapter la rubrique, Anna Nozza du Centre de coordination des essais cliniques de Montréal pour le calcul de la taille d'échantillon, l'équipe du Centre de simulation et les enseignants de la Faculté des sciences infirmières de l'Université de Montréal pour le soutien au recrutement, ainsi que tous les participants à l'étude.

## Références

1. Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care* 2004;13:i2-10.
2. Decker S, Fey M, Sideras S, Caballero S, Rockstraw L, Boese T, *et al.* Standards of best practice: simulation standard VI: the debriefing process. *Clin Simul Nurs* 2013;9:S26-9.
3. Gardner R. Introduction to debriefing. *Semin Perinatol* 2013;37:166-74.
4. Tannenbaum SI, Cerasoli CP. Do team and individual debriefs enhance performance? A meta-analysis. *Hum Factors* 2013;55:231-45.

5. Savoldelli GL, Naik VN, Park J, Joo HS, Chow R, Hamstra SJ. Value of debriefing during simulated crisis management: oral versus video-assisted oral feedback. *Anesthesiology* 2006;105:279-85.
6. Shinnick MA, Woo M, Horwich TB, Steadman R. Debriefing: the most important component in simulation? *Clin Simul Nurs* 2011;7:e105-11.
7. Cant RP, Cooper SJ. The benefits of debriefing as formative feedback in nurse education. *Aust J Adv Nurs* 2011;29:37-47.
8. Dreifuerst KT. The essentials of debriefing in simulation learning: a concept analysis. *Nurs Educ Perspect* 2009;30:109-14.
9. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc* 2007;2:115-25.
10. Dismukes RK, McDonnell LK, Jobe KK, Smith GM. What is facilitation and why use it? In: Dismukes RK, Smith GM (eds). *Facilitation and debriefing in aviation training and operations*. Ashgate: Aldershot (UK), 2000, p. 1-12.
11. Society for simulation in healthcare. Accreditation standards. 2016 [On-line]. Disponible sur : <https://www.ssih.org/Credentialing/Accreditation/Full-Accreditation>.
12. Royal College of Physicians and Surgeons of Canada. Simulation program accreditation: Accreditation standards. 2019 [On-line]. Disponible sur : <http://www.royalcollege.ca/rcsite/documents/continuing-professional-development/accreditation-simulation-programs-e.pdf>.
13. Cheng A, Grant V, Dieckmann P, Arora S, Robinson T, Eppich W. Faculty development for simulation programs: five issues for the future of debriefing training. *Simul Healthc* 2015;10:217-22.
14. Cheng A, Morse KJ, Rudolph J, Arab AA, Runnacles J, Eppich W. Learner-centered debriefing for health care simulation education: lessons for faculty development. *Simul Healthc* 2016;11:32-40.
15. Dismukes RK, Gaba DM, Howard SK. So many roads: facilitated debriefing in healthcare. *Simul Healthc* 2006;1:23-5.
16. LaFond CM, Blood A. Targeted simulation instructor course for nursing professional development specialists. *J Nurses Prof Dev* 2016;32:284-93.
17. Paige JT, Arora S, Fernandez G, Seymour N. Debriefing 101: training faculty to promote learning in simulation-based training. *Am J Surg* 2015;209:126-31.
18. Topping A, Boje RB, Rekola L, Hartvigsen T, Prescott S, Bland A, *et al.* Towards identifying nurse educator competencies required for simulation-based learning: a systemised rapid review and synthesis. *Nurse Educ Today* 2015;35:1108-13.
19. Nehring WM, Wexler T, Hughes F, Greenwell A. Faculty development for the use of high-fidelity patient simulation: a systematic review. *Int J Health Sci Educ* 2013;1:Art. 4.
20. Hallmark BF. Faculty development in simulation education. *Nurs Clin North Am* 2015;50:389-97.
21. Brett-Fleegler M, Rudolph J, Eppich W, Monuteaux M, Fleegler E, Cheng A, *et al.* Debriefing assessment for simulation in healthcare: development and psychometric properties. *Simul Healthc* 2012;7:288-94.
22. Arora S, Ahmed M, Paige J, Nestel D, Runnacles J, Hull L, *et al.* Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Ann Surg* 2012;256:982-88.
23. Feeney EJ. Quality feedback: the essential ingredient for teacher success. *The Clearing House* 2010;80:191-98.
24. Jonsson A, Svingby G. The use of scoring rubrics: reliability, validity and educational consequences. *Educ Res Rev* 2007;2:130-44.
25. Sousa VD, Rojjanasirrat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract* 2011;17:268-74.
26. Keeney S, Hasson F, McKenna H. *The delphi technique in nursing and health research*. West Sussex: Wiley-Blackwell, 2011.
27. Polit DF, Beck CT, Owen SV. Is the cvi an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007;30:459-67.
28. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-10.
29. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Pract Assess Res Eval* 2004;9:1-11.
30. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Med Teach* 2012;34:787-91.
31. Ahmed M, Sevdalis N, Paige J, Paragi-Gururaja R, Nestel D, Arora S. Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *Am J Surg* 2012;203:523-29.

**Citation de l'article** : Lavoie P., Deschênes M.-F., Brien L.-A., Radermaker M., Boyer L. Évaluation objective structurée de l'animation du débriefing (ÉOSAD) : traduction, adaptation et validation d'une rubrique. *Pédagogie Médicale* 2019;20:91-100