

Fidélité d'une grille d'évaluation de sites Internet de formation médicale

Reliability of an assessment checklist for medical education websites

Diane POIRIER¹, Marie-Françoise MÉGIE¹, Cristine LAMOUREUX¹, Johanne BLAIS¹

¹ Médecins francophones du Canada, Montréal, Canada

Manuscrit soumis à la rédaction le 13 septembre 2012 ; commentaires éditoriaux formulés aux auteurs le 21 avril et le 12 août 2013 ; accepté pour publication le 17 août 2013

Mots-clés

Développement professionnel médical continu ; évaluation ; grille ; sites Internet de formation médicale

Résumé – Contexte : La quantité de programmes de formation médicale disponibles sur Internet soulève un questionnement quant à l'uniformité de la qualité de l'information, la crédibilité et les sources du contenu médical retrouvé sur ces sites. **But :** Élaborer puis évaluer la fidélité d'une grille d'évaluation de sites Internet de formation médicale avec l'objectif d'offrir à la communauté médicale francophone internationale un outil utile, fidèle et innovant. **Méthode :** Une grille d'évaluation comportant 54 indicateurs groupés en cinq catégories a été élaborée. La fidélité du résultat global de la grille a été quantifiée en comparant les résultats obtenus par sept observateurs ayant évalué 16 mêmes sites Internet à deux reprises. La seconde évaluation des sites a eu lieu un an après la première afin de minimiser le risque que les observateurs se rappellent leur évaluation initiale. Ce plan de collecte de données a permis d'évaluer les composantes suivantes de la fidélité : la variabilité intra-observateurs, la variabilité inter-observateurs et l'interaction sites-observateurs. La fidélité du résultat global de la grille a été mesurée par le coefficient de corrélation intra-classe. **Résultats :** Le coefficient de corrélation intra-classe de la grille obtenu était de 0,684, ce qui correspond à une bonne fidélité. La variabilité intra-observateurs (répétabilité) a été la plus grande source de variation attribuable à la méthode de mesure avec 21 % de la variation totale. Une autre mesure de performance a aussi été utilisée : le rapport de discrimination. Le rapport de discrimination estime le nombre de catégories différentes pouvant être distinguées par un système de mesure. Le rapport de discrimination obtenu est de 2,3 ce qui indique que la grille permet de séparer les sites observés en deux catégories : faible et bien. **Conclusion :** La grille d'évaluation des sites Internet de formation médicale, par sa bonne fidélité, est un outil intéressant pour quantifier la qualité de ces sites. L'analyse démontre que la meilleure piste d'amélioration de la fidélité de la grille serait d'en améliorer la répétabilité (variabilité intra-observateurs) mais la consistance interne pourrait aussi être étudiée et optimisée.

Keywords

Continuing professional development in medicine;

Abstract – Context: The mass of available medical training programs on the Internet raises questions about the uniformity of the quality of information, the reliability and the sources of the medical information found on these sites. **Objectives:** Developing and then assessing the accuracy of an assessment checklist of medical education websites in view of providing to the international French-speaking medical community a useful, reliable and innovative tool.

evaluation tool;
checklist; medical
education websites

Method: An assessment checklist of 54 indicators grouped into five categories was developed. The accuracy of the checklist's overall score was compared using the results of seven observers who evaluated the same 16 websites twice. The second evaluation of the sites was carried out one year later one to minimize the risk of observers remembering their original evaluation. The data collection plan helped measure the following accuracy elements: intra-observer variability, inter-observer variability and site-observer interaction. The accuracy of the checklist score was measured by the intra-class correlation coefficient. **Results:** The resulting intra-class correlation coefficient was 0.684, consistent with a good degree of accuracy. Intra-observer variability (repeatability) was the largest source of variation attributable to the measuring method with 21% of total variation. Another performance measure was also used: the discrimination ratio that considers the number of different categories that can be recognized by a measurement system. The resulting ratio was 2.3, indicating that the checklist helps divide the sites identified in two categories: weak and average. **Conclusion:** Due to its high degree of accuracy, the assessment checklist of medical education websites is an interesting tool to assess the quality of these sites. The analysis clearly demonstrates that the best avenue in improving the reliability of the checklist would be to improve repeatability (intra-observer variability).

Introduction

Au cours des quinze dernières années, les programmes de formation médicale se sont multipliés sur Internet. La diversité des sites et de leurs origines complique le choix de l'utilisateur qui souhaite les utiliser comme ressources pour des apprentissages. D'un côté, plusieurs questions se posent quant à la qualité de l'information et à la crédibilité de ces sites en tant que sources d'informations médicales. De l'autre, les caractéristiques liées à la conception technique et à l'interactivité de ces sites sont des éléments importants qui conditionnent leur attractivité pour les professionnels de la santé, soumis à une offre de plus en plus importante de sites référencés, qui font parfois l'objet d'une promotion publicitaire importante.

Une revue systématique de la littérature consacrée à cette problématique, effectuée par les membres du comité de la recherche en formation professionnelle continue (FPC) de l'association « Médecins francophones du Canada », à partir d'une interrogation des bases documentaires Medline, MdConsult et Google, complétée par une recherche additionnelle à partir des références des articles retenus, a permis de colliger quelques articles traitant de la qualité de l'information médicale sur Internet. Toutefois, aucune publication dédiée à l'évaluation de sites de formation médicale n'a été trouvée. Cette revue de la

littérature a ainsi confirmé qu'il n'existait pas de grille connue qui puisse permettre, sur la base d'une analyse critériée et d'une mesure quantitative, d'évaluer et de comparer les sites de formation médicale.

La mise à disposition de la communauté médicale francophone internationale d'un outil fiable et innovant d'évaluation en langue française s'est donc avérée être une priorité pour les membres de ce comité, le but étant de faciliter la sélection de sites crédibles par les étudiants en médecine et les professionnels de la santé.

L'objectif de ce travail est de rapporter l'élaboration puis la validation métrologique de la fidélité d'une grille d'évaluation de sites Internet de formation médicale continue.

Matériel et méthodes

Construction de la grille d'évaluation et sélection de sites Internet éligibles à l'évaluation

La grille d'évaluation a été élaborée par les membres du comité de la recherche en FPC de « Médecins francophones du Canada » à partir de plusieurs sources : – deux études québécoises portant sur l'évaluation de sites d'information médicale sur Internet^[1-5] ;

- la littérature générale en matière de formation sur Internet et d'information en santé [2,6,7,8,9,13] ;
- les lignes directrices du développement professionnel continu (DPC) du Conseil québécois de développement professionnel continu des médecins (CQDPCM)^[10] ;
- l'expérience personnelle en DPC des membres du comité de la recherche en FPC de Médecins francophones du Canada.

Choix des catégories d'informations et des critères à documenter

Les catégories d'informations à documenter et les critères ont été choisis en s'inspirant des travaux du *Health Information Technology Institute* (HITI), publiés il y a plus de 15 ans, portant sur les critères d'évaluation de la qualité de l'information en santé sur internet^[8], du groupe de Rouen^[6] et des travaux de Bazin^[1] et de Frémont^[5]. Les critères jugés pertinents ont été adaptés et intégrés à la grille (tableau II). Ces critères concernaient d'une part, le contenu informationnel et sa crédibilité et, d'autre part, la présentation et l'interactivité des sites.

Le barème élaboré a attribué le pourcentage de points le plus élevé aux catégories « contenu » et « crédibilité », soit respectivement 42 % et 28 % du score total. Ce choix a pour but de prioriser la qualité et la crédibilité de l'information, éléments essentiels d'une bonne formation en ligne. Bazin^[1] et Frémont^[5] avaient pour leur part attribué 60 % des points à la catégorie « contenu » dans leur outil d'évaluation de sites d'information sur internet, la catégorie « crédibilité » n'ayant pas été retenue dans leurs travaux. La grille actuelle intègre cette catégorie jugée essentielle pour évaluer des sites de formation médicale.

Les critères retenus pour la catégorie « contenu » de la grille d'évaluation des sites de formation médicale diffèrent passablement de ceux retenus par Bazin^[1] et Frémont^[5]. En effet, l'exactitude des énoncés, la référence à des niveaux de preuve pour qualifier les données probantes et la mention d'éventuelles controverses sont des éléments majeurs dans la grille d'évaluation des sites de formation médicale, alors qu'ils n'avaient pas été retenus dans les grilles d'évaluation des sites d'information médicale. Pour

juger de la qualité d'un site de formation médicale, de tels indicateurs doivent permettre d'évaluer la rigueur et la validité de l'information.

Les critères retenus pour la catégorie « crédibilité » concernent : l'identification des objectifs de la formation, la pertinence de la formation, la mention de la source de l'information et du financement. Ces critères ont été jugés importants par le groupe de travail et ainsi intégrés à la grille.

Sélection de sites Internet éligibles à l'évaluation

Afin d'évaluer la grille, 55 sites Internet ont été identifiés par un sondage auprès de médecins résidents francophones. De ces 55 sites, 10 ont été éliminés car ils ne satisfaisaient pas un ou plusieurs des critères suivants : disponibilité et fonctionnalité ; accessibilité gratuite ; offre d'un programme de formation médicale et non seulement d'information ; fonctionnement asynchrone du site.

Des tests préalables ont ensuite été effectués à partir de trois sites Internet ; ils ont permis d'apporter des corrections à la grille.

Description de la grille

La grille se compose de cinq catégories d'informations, à documenter respectivement à l'aide de plusieurs critères, ces derniers étant eux-mêmes renseignés par plusieurs indicateurs. Chacune de ces catégories comporte entre quatre et 12 critères pour un total de 32.

Chacun des 32 critères est mesuré à l'aide de 1 à 4 indicateurs, pour un total de 54 indicateurs. Chacun des 54 indicateurs a été pondéré de 0, 5 à 5 points selon son importance. L'importance de chaque indicateur a été déterminée selon l'expérience des concepteurs de la grille et en s'inspirant du Net Scoring^[6] et de Bazin^[1]. Le tableau I présente le détail des cinq catégories de la grille et le tableau II les critères associés à chaque catégorie.

La sélection des indicateurs de chaque critère et leur pondération dans le score ont été faites par les auteurs en s'appuyant sur une étude de la littérature

Tableau I. Catégories de la grille d'évaluation et détails selon le nombre de critères, d'indicateurs et de points.

Catégorie	Détail
Crédibilité	6 critères composés d'un total de 14 indicateurs pour 28 points
Contenu	12 critères composés d'un total de 15 indicateurs pour 42 points
Conception technique	4 critères composés d'un total de 11 indicateurs pour 10 points
Interactivité	5 critères composés d'un total de 8 indicateurs pour 10 points
Éléments spécifiques aux modules	5 critères composés d'un total de 6 indicateurs pour 10 points

Tableau II. Critères associés aux cinq catégories de la grille d'évaluation des sites de formation médicale.

Catégorie	Critères
Crédibilité	objectifs, source, contexte, actualisation, pertinence/utilité, procédure éditoriale et choix de l'information
Contenu	Exactitude, niveaux de preuve, citation des sources originales, controverses, quantité d'information, mise à jour, réalité médicale, information à saisir, langue, gratuité ou non, outils d'aide à la décision, accréditation
Conception technique	Accessibilité, lisibilité, moteur de recherche interne, qualité du média
Interactivité	Retour d'information, forum, personnalisation interactive, hyperliens, évaluation par le visiteur
Éléments spécifiques aux modules	Calendrier et durée des activités, objectifs cohérents avec le format de l'activité, clarté, stratégies pédagogiques diversifiées, plan de développement professionnel

concernée, en vérifiant la légitimité de chaque indicateur à mesurer de façon valide les critères choisis. La pondération a été faite de façon à favoriser la catégorie « contenu » pour privilégier la qualité et la crédibilité de l'information.

Le score global généré par l'application de la grille résulte de l'addition du score élémentaire affecté à chacun des indicateurs selon leur pondération. Sa valeur peut donc varier entre 0 et 100.

Protocole de l'étude de fidélité de la grille d'évaluation

L'étude de fidélité de la grille a cherché à évaluer deux aspects de la fidélité du résultat global : la fidélité intra-observateurs (appelée aussi répétabilité), d'une part, qui évalue la concordance des évaluations répétées d'un même observateur sur un même site et la reproductibilité, d'autre part, qui inclut la fidélité inter-observateurs et l'interaction sites-observateurs, qui évalue la concordance des réponses entre des observateurs différents. Comme mentionné précédemment, la

fidélité intra-observateurs est évaluée en vérifiant que chaque observateur évalue au moins une seconde fois de façon concordante le même site Internet. Par contre, il est important de minimiser les risques que l'observateur, lors de la seconde évaluation du même site, se rappelle les résultats de sa première évaluation. C'est pourquoi l'étude s'est déroulée en deux phases (1 et 2) séparées par une période d'un an. À chacune des phases, des sites Internet communs ont été évalués par un groupe de sept observateurs, tous médecins résidents de première année (R1) de l'unité de médecine familiale Richelieu-Yamaska, affiliée à l'Université de Sherbrooke. Ces observateurs sont de sexe féminin et leur âge variait de 24 à 30 ans au début de l'étude. Les sept observateurs ont été identifiés comme suit dans la base de données : 1-EB, 2-ACR, 3-AC, 4-CG, 5-MCC, 6-ML, 7-JB.

Lors de la première phase, 20 sites Internet ont été choisis au hasard parmi les 45 sites répondant aux critères d'inclusion. Parmi ces 20 sites, quatre ont dû être éliminés (les sites 9, 14, 15 et 19) parce qu'ils ont été jugés inapplicables ou parce que certains de ces sites n'ont pas été évalués par l'un des observateurs.

Lors de la seconde phase de l'étude, les 16 sites évalués lors de la première phase ont été à nouveau évalués par l'ensemble des observateurs. Entre les deux phases, la grille d'évaluation a été ajustée. À la suite des analyses des réponses obtenues lors de la première phase, certaines questions ont été modifiées, afin d'en clarifier la signification et l'interprétation. Chaque catégorie de questions a fait l'objet de quelques ajustements mineurs.

Les résultats des sept observateurs ont ensuite été compilés une première fois par un membre du comité de la recherche en FPC de « Médecins francophones du Canada », puis par une personne indépendante. Cette double vérification a permis d'identifier et de corriger quelques erreurs de saisie de résultats.

Objectif de l'analyse de fidélité de la grille

L'objectif principal de l'analyse de fidélité de la grille est de déterminer la fidélité du résultat global en termes de variabilité intra-observateurs, variabilité inter-observateurs et interaction sites-observateurs.

Description des analyses statistiques

Afin de pouvoir évaluer la fidélité intra-observateurs, chaque site a été évalué à deux reprises à un an d'inter-valle. Une analyse de la variance (ANOVA) à effet fixe a été effectuée sur les différences entre le résultat global des deux années. Les facteurs sites et observateurs ont été inclus dans l'ANOVA (le facteur interaction sites-observateurs n'a pu être inclus puisque le nombre de degrés de liberté ne le permet pas).

Statistiquement, la fidélité de la grille d'évaluation a été évaluée par le coefficient de corrélation intra-classe (*CCI*). Le *CCI* correspond au pourcentage de la variabilité observée dans les résultats, attribuable à de vraies différences entre les sites. Un pourcentage de *CCI* élevé correspond donc à un instrument de mesure dont les variations intra et inter-observateurs, ainsi que l'interaction sites-observateurs, sont faibles comparativement à la variation observable entre les sites. Le *CCI* est calculé par un ratio des variances obtenues par une ANOVA à effet aléatoire. Une ANOVA à effet aléatoire est utilisée puisque les sites ainsi que les observateurs ne sont

qu'un échantillon aléatoire parmi les sites et évaluateurs possibles. L'ANOVA à effet aléatoire permet de séparer la variation totale en quatre sources : la variabilité inter-sites, la variabilité intra-observateurs, la variabilité inter-observateurs et l'interaction sites-observateurs. Le *CCI* est obtenu en divisant la variance totale de la mesure (somme de la variance intra-observateurs, inter-observateurs et interaction sites-observateurs) par la variance totale observée lors de l'étude. L'interprétation du *CCI* selon Cicchetti et Sparrow^[4] et Cicchetti^[3] permet de qualifier la fidélité de la façon suivante : $CCI < 0,40$: fidélité faible ; $0,40 < CCI < 0,59$: fidélité passable ; $0,60 < CCI < 0,74$: fidélité bonne ; $CCI > 0,74$: fidélité excellente.

Une dernière mesure moins usuelle mais intéressante est le rapport de discrimination^[12]. Ce chiffre représente le nombre de catégories différentes pouvant être discriminées par l'outil de mesure. Par exemple, un rapport de discrimination de 2 signifie qu'un outil de mesure est capable de distinguer deux catégories, bon et mauvais, par exemple. Le rapport de discrimination (*D*) est calculé à partir des résultats de l'ANOVA à effet aléatoire. Il y a donc un lien entre le *CCI* et le rapport de discrimination. Le rapport de discrimination *D* est calculé à l'aide de l'équation suivante, où *CCI* est le coefficient de corrélation intra-classe :

$$D = \sqrt{2 \left(\frac{\text{Variance inter sites}}{\text{Variance totale} - \text{Variance inter sites}} \right) + 1}$$

$$= \sqrt{2 \left(\frac{CCI}{1 - CCI} \right) + 1}.$$

Le *CCI* et le rapport de discrimination ont été calculés à partir du résultat global de chaque site.

Résultats

L'analyse statistique des résultats a comparé initialement les résultats de la phase 1 à ceux de la phase 2, afin de valider si les différences observées entre les deux phases étaient aléatoires ou s'il y avait un biais observable, ou un impact des opérateurs ou des sites sur les différences. Cette étape a permis de mieux

Tableau III. Analyse de variance –ANOVA- à effet fixe sur la différence entre les résultats globaux de la phase 2 et ceux de la phase 1.

Composante	Degrés de liberté	Carrés moyens	Ratio <i>F</i>	Valeur- <i>p</i>
Sites	15	67,5	1,14	0,33
Observateurs	6	57,2	0,975	0,45
Erreur	90	59		
Total	111			

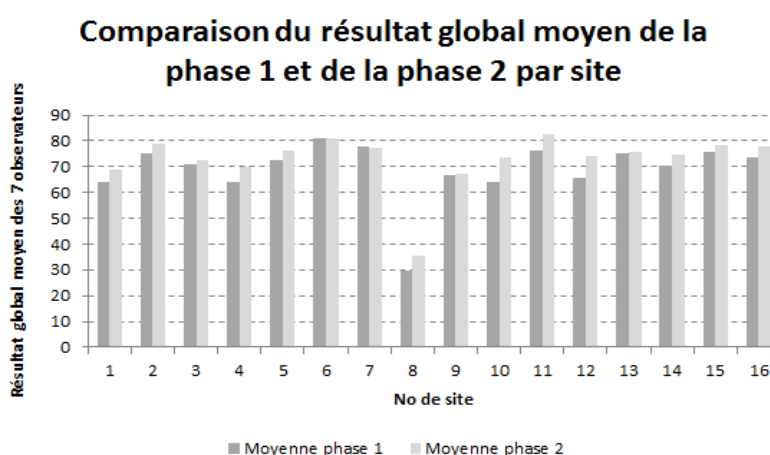


Fig. 1. Comparaison du résultat global moyen de la phase 1 et de la phase 2 par site.

comprendre ce qui pouvait affecter l'estimation de la variation provenant de la répétabilité (aussi appelée différence intra-observateurs). La seconde analyse a consisté en une ANOVA à effet aléatoire sur les résultats globaux. C'est à partir de celle-ci que le *CCI* et le rapport de discrimination ont été obtenus.

Différence entre la phase 1 et la phase 2

Une ANOVA à effet fixe a été effectuée sur les différences entre les résultats globaux de la phase 2 et ceux de la phase 1. Le modèle utilisé pour l'ANOVA comprenait les facteurs sites et observateurs. Les résultats de l'ANOVA à effet fixe sur la différence entre les résultats globaux de la phase 2 et ceux de la phase 1 sont présentés au tableau III.

La différence moyenne entre les deux phases est estimée à 3,9 et elle est statistiquement significative (valeur de la statistique *t* : 5,34 avec 111 degrés de liberté, soit *p* inférieur à 0,001). Cette différence

moyenne est obtenue en soustrayant le résultat global de la phase 1 de celui de la phase 2 pour chaque site et chaque observateur (un total de 112 valeurs correspondant à 16 sites et 7 observateurs) puis en calculant la moyenne de ces 112 différences.

La figure 1 présente, pour chaque site, la moyenne des résultats des sept observateurs pour la phase 1 et la phase 2.

Fidélité du résultat global de la grille

ANOVA à effet aléatoire

Une ANOVA à effet aléatoire a été effectuée sur l'ensemble des résultats globaux des deux phases.

Cette analyse de variance a permis d'estimer la variance attribuable à chaque source (inter-sites, inter-observateurs, interaction sites-observateurs et intra-observateurs) et ainsi obtenir le *CCI* (correspondant au pourcentage de variation inter-sites) tel que présenté au tableau IV.

Tableau IV. Composantes de la variance totale pour les phases 1 et 2 combinées.

Composante	Variance estimée	Pourcentage du total	Valeur- <i>p</i>
Inter-sites	121,804	68,4	< 0,0001
Reproductibilité de la mesure			
Inter-observateurs	5,10981	2,9	0,0036
Interaction sites-observateurs	13,8275	7,7	0,0028
Répétabilité de la mesure			
Intra-observateurs	37,3929	21,0	
Total	178,134	100	

Tableau V. Composantes de la variance totale pour les phases 1 et 2 combinées pour les résultats globaux ajustés.

Composante	Variance estimée	Pourcentage du total	Valeur- <i>p</i>
Inter-sites	121,804	69,9	< 0,0001
Reproductibilité de la mesure			
Inter-observateurs	5,1098	2,9	0,0036
Interaction sites-observateurs	17,651	10,1	< 0,0001
Répétabilité de la mesure			
Intra-observateurs	29,746	17,1	
Total	178,1342	100	

Il est possible de calculer un intervalle de confiance sur le CCI en utilisant l'approximation de Leiva et Graybill^[11]. L'intervalle de confiance de niveau 95 % pour le CCI obtenu est [0,478 ; 0,846]. Enfin, le rapport de discrimination de la grille est de 2,3.

ANOVA à effet aléatoire après correction pour la différence moyenne observée entre les deux phases

Il a été identifié précédemment qu'une différence moyenne de 3,9 dans le résultat global avait été observée entre la phase 2 et la phase 1. La valeur 3,9 a été ajoutée aux résultats globaux de la phase 2 afin de quantifier l'impact de cette différence sur le CCI. Ainsi, le biais de 3,9 entre les deux études a été mathématiquement éliminé. Une ANOVA à effet aléatoire a été effectuée sur les résultats ajustés. Le tableau V contient les composantes de la variance obtenue par cette ANOVA.

Une analyse graphique présentée à la figure 2 (le graphique multi-vari) présente les résultats obtenus par les observateurs selon les sites. La droite verticale présente pour chaque site représente la variation intra-sites. Les droites reliant la moyenne des sites

représentent la variation inter-sites. La variabilité inter-observateurs est visualisable par les lignes horizontales associées à chaque observateur. En examinant les lignes horizontales, il semblerait que deux observateurs aient obtenu des résultats qui sont, en moyenne, plus élevés (3-AC et 7-JB) et que deux observateurs ont obtenu des résultats qui sont, en moyenne, plus bas (1-EB et 6-ML). De plus, la variabilité provenant de l'interaction pourrait s'expliquer principalement par les moyennes très différentes entre les observateurs pour le site 8.

Discussion

Interprétation des résultats

Différences entre la phase 2 et la phase 1

Il existe une différence moyenne statistiquement significative -de 3,9 en moyenne- entre les résultats de la seconde phase et ceux de la première phase. Cette différence n'est pas significativement différente selon les observateurs ou les sites. Une telle différence de moyenne entre les deux phases pourrait ainsi s'expliquer par le fait que les sites se sont améliorés dans

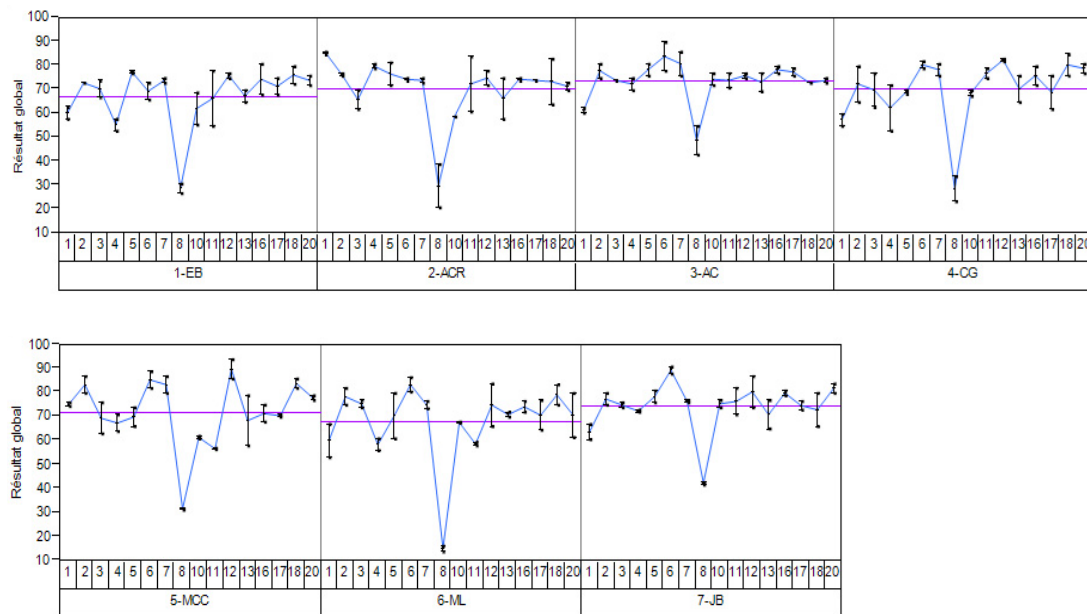


Fig. 2. Graphique multi-vari des résultats globaux en fonction des sites et des observateurs.

l'intervalle d'un an entre les deux phases, et/ou par le changement de la formulation de certaines questions entre les deux phases, et/ou par une meilleure compréhension des questions par les observateurs à la phase 2, avec en conséquence une meilleure aptitude à identifier la présence d'éléments du site à évaluer, et/ou par un autre facteur inconnu à ce moment.

Fidélité du résultat global de la grille

Nos résultats montrent que les variabilités inter-sites, inter-observateurs et liée à l'interaction sites-observateurs sont statistiquement significatives.

La grille d'évaluation des sites médicaux a un CCI de 0,684 correspondant à une bonne fidélité selon Cicchetti³. L'intervalle de confiance de niveau 95 % pour le CCI est [0,478 ; 0,846] ce qui signifie que si l'échantillon de sites et observateurs est représentatif de la population, le « vrai » CCI devrait se retrouver dans cet intervalle avec un niveau de confiance de 95 %. La précision du CCI fait en sorte que la fidélité de la grille pourrait en fait être de passable à excellente^[3,4].

Le rapport de discrimination de 2,3 indique que le nombre de catégories différentes de sites pouvant être identifiées parmi les sites observés lors de l'étude est d'environ deux. Ceci revient à dire que les sites observés dans cette étude peuvent être classés, par exemple, comme étant dans la catégorie « bien » ou dans la catégorie « faible ».

Le CCI obtenu sur les résultats ajustés est très légèrement supérieur à celui obtenu sur les résultats non ajustés (0,699 comparativement à 0,684). Par conséquent, la différence moyenne de 3,9 observée entre les deux phases n'affecte pas significativement les conclusions relatives à la fidélité de la grille. Outre la variabilité inter-sites, c'est encore la variabilité intra-observateurs (répétabilité) qui contribue le plus à la variabilité avec 17,1 % de contribution.

Limites de l'étude

L'étude de la fidélité des résultats globaux de la grille a été effectuée à l'aide de sept observateurs qui étaient tous des résidents de première année (R1) en médecine familiale, de sexe féminin et dont l'âge

variait de 24 à 30 ans au début de l'étude. La généralisation des résultats quant à l'impact des observateurs sur la fidélité est donc limitée puisque cet échantillon d'observateurs n'est pas représentatif de tous les utilisateurs potentiels de la grille.

Les résultats démontrent que la grille d'évaluation des sites Internet de formation médicale présente une bonne fidélité dans un contexte où les sept observateurs n'ont reçu aucune formation sur la grille visant à en standardiser l'utilisation. Le biais de 3,9 points entre les deux phases est surprenant mais n'affecte pas substantiellement le *CCI* obtenu. L'analyse démontre clairement que la meilleure piste d'amélioration de la fidélité de la grille serait d'en améliorer la répétabilité (variation intra-observateurs).

Il faut noter que la valeur du *CCI* est un estimé dont la précision est fonction, entre autres, de la taille de l'échantillon. Dans le cadre de l'actuelle étude, 16 sites, 7 observateurs et 2 mesures (phase 1 et 2) ont permis de collecter 224 observations. L'intervalle de confiance de niveau 95 % du *CCI* a démontré que la fidélité pourrait en fait être de passable à excellente. Une étude avec un plus grand nombre d'observations pourrait permettre d'obtenir un estimé plus précis du *CCI*.

Par ailleurs, nous n'avons pas procédé à l'analyse de la fidélité au regard de la dimension « consistance interne » des différents éléments de la grille. Une telle étude, qui pourrait être effectuée dans une prochaine phase de recherche, pourrait permettre d'identifier des pistes d'amélioration du questionnaire. Cependant, la fidélité obtenue à partir des résultats globaux indiquent que la grille, bien qu'elle puisse probablement être améliorée, démontre un potentiel intéressant dans sa capacité à évaluer la qualité des sites de formation médicale.

Une application envisagée de la grille est de l'utiliser comme outil pour sélectionner certains sites selon leur contenu et leur crédibilité lors de formations auprès d'étudiants en médecine et autres professionnels de la santé. De plus, certains sites pourraient être recommandés en se basant surtout sur l'interactivité et la conception technique (par exemple, l'utilisation de vidéos pour l'apprentissage de certaines techniques ou simulation). L'étude effectuée dans le cadre de ce travail permet de conclure que la grille

peut discriminer deux catégories de sites dans la plage de variation observée. Toutefois, le résultat global permettant de catégoriser un site comme « bien » ou « faible » n'a pas formellement été défini. Selon l'expérience des concepteurs de la grille, une valeur de 60 % serait suggérée afin de séparer les sites « faibles » des sites « bien ». Il n'est pas tout à fait clair, pour un utilisateur de la grille, de juger si un résultat obtenu pour un site est une indication que le site n'est pas d'un niveau de qualité suffisant puisque la décision de consulter un site Internet de formation dépend du but recherché par l'utilisateur.

Un aspect pouvant freiner l'utilisation de la grille est le temps requis pour répondre à tous les indicateurs. La construction d'une version abrégée (version courte) est considérée.

Finalement, une autre application possible de cette grille serait l'évaluation de plusieurs sites de formation médicale sur Internet par un groupe témoin de personnes formées à son utilisation. Le résultat moyen obtenu permettrait possiblement de discriminer entre plus de deux catégories. L'impact sur la fidélité de la grille d'évaluation qu'aurait le recours à un groupe témoin constitué de personnes spécialement formées pourrait être évalué lors d'une prochaine étude. Un tel groupe pourrait permettre de cibler des sites Internet ayant obtenu un résultat moyen acceptable.

Conclusion

La présente étude visait à évaluer la fidélité des résultats globaux de la grille d'évaluation de sites Internet de formation médicale selon les aspects suivants : variabilité inter-sites, variabilité inter-observateurs, interaction sites-observateurs.

Les analyses statistiques des données obtenues à partir de l'utilisation de la grille pour l'évaluation des sites de formation médicale dans le cadre de deux séries de mesures séparées par une période d'un an ont permis de vérifier que la fidélité de la grille, au regard de ces trois dimensions, est bonne, qualifiable de passable à excellente selon l'intervalle de confiance de niveau 95 % sur le *CCI*.

L'analyse statistique a permis de séparer la variabilité provenant de la reproductivité de celle

provenant de la répétabilité. La répétabilité (variabilité intra-sites/observateurs) contribue davantage que la reproductibilité à la variabilité totale observée.

Une étude approfondie de la consistance interne des éléments de la grille pourrait permettre de mieux comprendre les causes possibles de discordance et d'apporter des améliorations de nature à mettre à disposition des utilisateurs un outil d'évaluation encore plus performant.

Contributions

Tous les auteurs ont participé à la conception de l'étude et du protocole de recherche, ainsi qu'à l'interprétation des résultats. Diane Poirier a participé au recueil des données et à l'analyse statistique, avec la collaboration de Martin Carignan, statisticien. Diane Poirier et Marie-Françoise Mégie ont rédigé la version initiale du manuscrit, avec la collaboration de Martin Carignan pour la partie statistique. Diane Poirier, Martin Carignan, Marie-Françoise Mégie et Johanne Blais ont collaboré pour la révision du manuscrit

Ce travail a également bénéficié des contributions de Martin Carignan, statisticien au groupe conseil en statistiques Différence, Mme Céline Monette et Mme Diane Bircher de Médecins francophones du Canada.

Remerciements

Les auteurs remercient les résidentes en médecine familiale de la promotion 2009-2011 de l'Unité de médecine familiale Richelieu-Yamaska.

Approbation éthique

Le projet a été approuvé par le comité d'éthique de la recherche du Centre de santé et de services sociaux (CSSS) Richelieu-Yamaska en date du 16 juin 2009.

Subvention

Le travail a bénéficié d'une bourse de recherche 2009 en développement professionnel continu de la

part du Conseil québécois de développement professionnel continu des médecins (CQDPCM) pour sa réalisation et sa diffusion.

Autres valorisations scientifiques

Le travail a fait l'objet d'une communication affichée présentée au 5^e Forum international francophone de pédagogie des sciences de la santé (FIFPSS) qui s'est tenu à Québec (Québec-Canada) du 16 au 18 juin 2010, d'une communication orale aux XVII^e Journées Universitaires Francophones de pédagogie des sciences de la santé de la CIDMEF, qui se sont tenues à Lille (France) du 5 au 8 avril 2008, ainsi que d'un article paru dans *La Lettre du DPC* (Volume 19, Numéro 3, Septembre 2010) du Conseil québécois de développement professionnel continu des médecins (CQDPCM)

Références

1. Bazin L. Élaboration d'une grille de sélection des sites médicaux. *Bulletin des Bibliothèques de France* (Paris) 1999;44:73-6.
2. Choules AP. The use of elearning in medical education: a review of the current situation. *Postgrad Med J* 2007;83:212-216.
3. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284-90.
4. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86:127-137.
5. Frémont P, Labrecque M, Légaré F, Baillargeon L, Misson L. Evaluation of medical web sites: Interobserver and intraobserver reliability of an evaluation tool. *Can Fam Physician* 2001;47:2270-8.
6. Groupe de Rouen. Net Scoring: Critères de qualité de l'information de santé sur l'internet 2001 [On-line] Disponible sur : <http://www.chu-rouen.fr/netscoring/>

7. Harris JM, Sklar BM, Amend RW, Novalis-Marine CJ. The growth, characteristics, and future of online CME. *J Contin Educ Health Prof* 2010;30:3-10.
8. Health Information Technology Institute. Quality of health information on the internet. 1999 [On-line] Disponible sur : <http://www.ahrq.gov/research/data/infoqual.html>
9. Horton W. *Designing Web-Based Training*. New York: Wiley, 2000.
10. Raïche P. Développement professionnel continu en ligne : crédits. *La Lettre du DPC* 2009;18:5.
11. Leiva RA, Graybill FA. Confidence intervals for variance components in the balanced two-way model with interaction. *Communications in Statistics: Simulation and Computation* 1986;15:301-22.
12. Wheeler DJ, Lyday RW. *Evaluating the Measurement Process (Second Edition)* Knoxville: SPC Press, 1989.
13. Zajaczek JEW, Götz F, Kupka T, Behrends M. eLearning in education and advanced training in neuroradiology: introduction of a web-based teaching and learning application, *Neuroradiology* 2006;48:640-646.

Correspondance et offprints : Diane Poirier, Hôpital Honoré-Mercier, 2750 Bd. Laframboise, J2S 4Y8 Saint-Hyacinthe Québec, Canada.
Mailto: dianepoirier2@cgocable.ca