
Les grilles d'observation pour évaluer les apprentissages

Dominique Houle
Louise Ménard
Professeures
Soins infirmiers

Robert Howe
Conseiller pédagogique
Collège Montmorency

L'évaluation des apprentissages en stage a toujours posé de nombreux défis aux enseignantes et aux enseignants. Il y a d'abord la préoccupation de l'impartialité. Le professeur évaluateur est souvent seul avec l'étudiant évalué. Le geste est éphémère et il en reste souvent peu de traces, surtout quand on évalue un processus ou une attitude. Ce qui paraît prioritaire ou inacceptable pour un enseignant peut parfois relever de son propre système de valeurs et altérer son jugement. Comment minimiser l'influence des biais personnels au moment de porter un jugement à la fin du stage ? Puis, il y a la question de l'équivalence entre les notes attribuées. Une performance moyenne pour un enseignant représente-t-elle une performance moyenne pour un autre ? La même note est-elle attribuée à deux stagiaires qui ont une performance équivalente ? Et, enfin, comment juger une prestation qui se situe tout juste à la limite de l'acceptable ou de l'inacceptable ? Comment s'assurer que le verdict de réussite ou d'échec est réellement justifié ?

Les étudiants remettent aussi souvent en question les façons d'évaluer les stages. Combien de fois rencontre-t-on des étudiants qui critiquent soit l'impartialité des jugements, soit l'équivalence des notes attribuées ? Cela donne occasionnellement droit à des rondes de négociation ou même à des situations de contestation agressives.

C'est justement suite à une situation de stage difficile à vivre par une de nos collègues en soins infirmiers et l'une de ses étudiantes que nous avons décidé de revoir encore notre grille d'évaluation. Nous étions deux enseignantes prêtes à relever le défi. Nous visions alors la création d'un outil présentant des critères suffisamment explicites pour permettre aux enseignantes de poser leur jugement en s'appuyant sur les mêmes observations et pour les aider à distinguer les différents niveaux d'une performance. Nous voulions que le processus d'évaluation soit rendu plus transparent tant pour les enseignantes que pour les étudiantes. Nous projetions également de faire participer activement nos collègues à l'élaboration de cette grille afin qu'elles se l'approprient et qu'elles la voient comme représentative de leur propre démarche. Finalement, nous souhaitions qu'elle soit utile aux étudiantes en leur fournissant des indications claires sur ce qui est attendu d'elles en stage. À nos yeux, cela favoriserait la prise en charge des apprentissages par les stagiaires. Nous avons donc entrepris, avec l'aide d'un conseiller pédagogique, de mettre au point et d'expérimenter une forme évoluée de grille d'observation, fondée sur une échelle descriptive, qui servirait à l'évaluation sommative pour le cours « Stage d'intégration » (180-620).

Une grille d'observation à échelle descriptive

La grille d'observation, aussi appelée grille d'appréciation, est un instrument de mesure qui permet de constater les particularités d'une action, d'un produit ou d'un processus¹. De façon générale, cette grille comporte deux volets caractéristiques : une liste d'éléments à observer et une façon d'enregistrer ce que nous

observons. Les éléments à observer, ce sont les critères sur lesquels nous désirons porter un jugement de valeur. Dans une grille d'observation, le critère n'est pas une phrase complète. Il est constitué des mots essentiels qui le décrivent et qui le distinguent d'un autre critère. Par exemple : la rigueur du respect professionnel ; la régularité et la fréquence de la collaboration avec les pairs et avec les membres de l'équipe soignante ; l'empathie ; etc.

Afin de pouvoir enregistrer ce que nous observons, il nous faut une échelle. Celle que nous avons retenue est dite « descriptive »². Elle est composée d'énoncés qui sont des phrases simples, mais complètes, avec sujet, verbe et complément, et qui servent à décrire un comportement.

Pour chaque critère, les énoncés descriptifs ont en commun la description de ce que le juge cherchera à observer, mais dépouillée de ce qui qualifie et nuance le comportement ou la performance. Il s'agit du comportement qui indique la présence du critère. Plusieurs indicateurs peuvent « signifier » la présence du critère. Pour le critère d'évaluation « Autonomie dans la rédaction des plans de soin », par exemple, le dénominateur commun à chacun des énoncés descriptifs peut être le suivant : « *Les quatre derniers jours de stage, l'étudiante rédige les plans de soins de ses clients.* » On voit tout de suite que ce dénominateur commun détermine déjà la tâche de l'étudiante pendant la portion de son stage qui fera l'objet de l'évaluation sommative. Ce même dénominateur commun constitue la cible de l'observation de l'enseignante.

Les énoncés descriptifs se distinguent cependant les uns des autres du fait que chacun détermine un niveau de performance particulier pour le critère considéré. Selon qu'on décrit un niveau insuffisant, minimal, moyen ou supérieur, on utilisera un adjectif, un adverbe, un nombre, une durée, etc. qui servent à caractériser la performance, à distinguer les performances entre elles et à guider l'observateur dans son jugement.

Par exemple, pour décrire un niveau minimalement acceptable, on pourra écrire : « *Les quatre derniers jours de stage, l'étudiante nécessite un encadrement assez fréquent de la part de la professeure lorsqu'elle rédige les plans de soins de ses clients* », alors que pour décrire un niveau inacceptable, on écrira : « *Les quatre derniers jours de stage, l'étudiante nécessite un encadrement fréquent de la part de la professeure lorsqu'elle rédige les plans de soins de ses clients* » et « *Les quatre derniers jours de stage, l'étudiante rédige tous les plans de soins de ses clients de façon autonome* » pour signifier un comportement jugé supérieur.

L'échelle descriptive aide l'observateur à toujours focaliser son attention, d'une étudiante à l'autre, sur un même indicateur de la compétence. Cette stabilité dans l'observation favorise la fidélité des résultats de l'évaluation. D'une étudiante à l'autre, l'observateur est constamment rappelé à l'ensemble des caractéristiques du comportement qu'il cherche à observer.

Si plus d'un observateur participe à l'observation et, *a fortiori*, si l'opération d'évaluation s'étale sur plusieurs jours, les échelles descriptives permettent aux observateurs d'avoir une compréhension univoque d'un critère et d'observer l'objet de l'évaluation de la même manière. Dans un tel cas, les échelles descriptives favorisent la stabilité (fidélité) inter-juges.

Mise au point d'une grille d'évaluation à échelle descriptive

La mise au point d'une échelle descriptive comporte deux grandes étapes : d'une part, le choix des critères et la rédaction des énoncés descriptifs, et, d'autre part, l'assemblage de la grille.

Choix des critères et rédaction des énoncés descriptifs

Pour cette étape, nous proposons une démarche qui peut grandement faciliter le travail en équipe.

- Énumérer tous les critères sur lesquels on veut faire une évaluation.
- S'assurer que l'énoncé du critère est clair et univoque pour tous.
- Faire travailler l'équipe sur un critère à la fois et l'amener à rédiger tous les énoncés descriptifs portant sur ce critère.
- Faire rédiger les énoncés descriptifs dans l'ordre suivant :
 1. le niveau de la compétence minimale ;
 2. le niveau de la compétence insuffisante ;
 3. le niveau de la compétence supérieure ;
 4. le niveau de la compétence moyenne.
- Permettre aux membres de l'équipe de négocier des ajustements aux libellés des énoncés, afin qu'ils puissent bien caractériser chacun des niveaux et être à l'aise avec les énoncés. Ces ajustements seront faits notamment en modifiant les adjectifs ou les adverbes, les nombres, la durée, etc.
- Pour faciliter la formulation d'un énoncé descriptif, aider les membres de l'équipe à se focaliser sur le sens (l'esprit) de chaque niveau de compétence en suggérant des questions comme :
 - Selon votre jugement expert, qu'est-ce qui caractérise la compétence minimale (ou insuffisante, ou moyenne, etc.) sur ce critère ?
 - Pour ce critère, quelle compétence minimale exigerez-vous d'une étudiante, au moment de cette évaluation sommative, pour que vous acceptiez de lui accorder une mention de réussite minimale ?
 - Pour ce critère, que ferait une étudiante pour que vous jugiez qu'elle mérite un échec (ou qu'elle a une compétence supérieure aux autres, etc.) ?

Tous les énoncés descriptifs des divers critères d'évaluation peuvent donner lieu à un document d'une certaine ampleur. Si une grille d'observation à échelle numérique simple peut habituellement tenir sur une seule page, une grille à échelle descriptive sera plus longue, s'étalant parfois sur quatre ou cinq pages.

Certains seront tentés d'économiser de l'espace ou du temps de lecture en faisant l'abréviation des énoncés descriptifs. Il n'est pas conseillé d'abrégier les énoncés parce qu'on risque de provoquer un effet pervers. Le lecteur, professeur-observateur, peut ainsi perdre de vue le sens du descriptif. Il risque de s'en tenir à des mots clés et il devient alors inutile

d'avoir rédigé des énoncés complets au départ. Plus on simplifiera les énoncés descriptifs, plus on se rapprochera des inconvénients de l'échelle numérique simple qui ne propose, comme guide de jugement, que des expressions comme « excellent », « bien », « mauvais », « insuffisant ». Ce qui est « excellent » pour un juge peut n'être que « très bien » pour un autre.

Par ailleurs, puisque les énoncés descriptifs des divers niveaux de jugement pour un critère sont morphologiquement semblables, il peut être tentant de souligner les mots qui les distinguent les uns des autres. Là aussi, cette pratique risque d'inciter l'utilisateur à s'en tenir

aux mots soulignés, dans le vif de l'action (« toujours », « souvent », « un sur deux », « la plupart du temps », etc.) et à perdre de vue l'esprit de l'énoncé complet.

Assemblage de la grille d'observation à échelle descriptive

Si la plupart des critères peuvent se juger sur une échelle en trois, quatre, cinq ou même six niveaux³, certains ne se jugent que de façon dichotomique : oui ou non ; présent ou absent ; vrai ou faux ; réussite ou échec. Par exemple, en soins infirmiers, un critère portant sur l'honnêteté ou sur la déclaration des erreurs de soins ne souffre pas de compromis : il n'y a pas de demi-réussite sur ces aspects. Dans un tel cas de mixité d'échelle, il est préférable de regrouper les critères en distinguant clairement ceux qui se mesurent sur une échelle en deux niveaux de ceux qui se mesurent sur une échelle à plusieurs niveaux. Il sera alors important que l'observateur se familiarise avec la grille et note bien que certains critères ont une échelle de jugement différente des autres. Les échelles doivent bien entendu être accompagnées de consignes claires quant aux conditions d'observation à des fins d'évaluation sommative. De plus, les règles selon lesquelles on arrivera à une note finale et à un verdict de succès ou d'échec doivent être définies. Ces règles, que nous appelons algorithme de décision, doivent accompagner la grille d'observation à échelle descriptive.

La grille d'observation en soins infirmiers

Les critères d'évaluation

Pour déterminer les critères, nous avons commencé notre démarche en examinant l'outil que nous utilisons déjà. Celui-ci avait été élaboré à partir des objectifs du stage en tenant compte des normes et des critères de compétence énoncés par notre corporation professionnelle, l'Ordre des Infirmières et Infirmiers du Québec. Pour développer

notre nouvel outil, nous avons décidé de prendre aussi en compte le récent document d'orientation de la pratique infirmière de l'OIIQ⁴, qui remplace le document à partir duquel notre grille avait été construite. Nous avons finalement retenu vingt et un critères pour évaluer le stage.

L'échelle de jugement

La rédaction des énoncés descriptifs s'est avérée beaucoup plus simple que prévu. Il nous a suffi d'imaginer à partir de quels comportements nous posions notre jugement. Par exemple, nous nous sommes rendu compte que, pour décider du niveau de motivation d'une étudiante, nous regardions dans quelle mesure elle montrait de l'intérêt pour assister à des démonstrations ou pour utiliser de nouvelles méthodes de soins. Nous avons décrit chacun des critères retenus en procédant de la même manière. Le plus souvent, nous débutons l'exercice par la formulation de la performance minimale. Puis nous décidons des nuances à considérer pour les niveaux de compétence insuffisant, supérieur et moyen. La description du niveau insuffisant a été sans doute la plus difficile car elle signifie l'échec au critère.

Dans certains cas, nous avons compris qu'il n'existait que deux niveaux de performance. Par exemple, une étudiante est honnête ou ne l'est pas. C'est ainsi que s'est introduite la notion de critère dichotomique. Nous avons aussi réalisé que certains critères n'étaient pas observables directement et qu'il fallait utiliser des moyens indirects pour les mesurer. Par exemple, il n'est pas possible d'assister à une communication de type relation d'aide sans la perturber ou même la compromettre. Nous avons donc pensé évaluer la capacité de l'étudiante à réfléchir de façon critique à cette communication de type thérapeutique.

Pour illustrer notre propos, nous présentons sur la page suivante un échantillon de notre grille.

L'algorithme de décision

Après consultation auprès des enseignantes du cours concerné, nous avons décidé d'octroyer une valeur égale à chacun des critères d'évaluation. Par ailleurs, puisque nous savions que la réussite de certains critères était indispensable pour faire foi de la performance minimale d'une finissante et de sa capacité à faire face aux exigences du marché du travail, nous avons également décidé que quinze des vingt et un critères évalués doivent être réussis de façon minimale pour que la note soit calculée. Un échec à l'un de ces quinze critères constitue un échec au stage. Le processus de calcul de la note, transparent et simple, s'effectue donc en deux étapes : il faut d'abord confirmer que l'étudiante réussit les critères obligatoires avant que les autres critères soient pris en considération dans le calcul de la note finale.

Nous avons bien entendu informé clairement les étudiantes de notre algorithme de décision et nous nous assurons, en distribuant à l'avance l'ensemble du document d'évaluation sommative, que les étudiantes comprennent bien cet algorithme tout autant que l'ensemble de la grille.

L'expérimentation

La première version de notre grille a été enrichie grâce aux commentaires d'autres enseignantes du département. Ainsi, nous pouvons dire qu'un consensus d'experts s'est établi.

Une première utilisation a été faite en milieu clinique au trimestre d'hiver 1997, dans une unité de soins intensifs intermédiaires, par une des conceptrices. Suite à cette étape, quelques éléments mineurs ont été éliminés de la grille puisqu'ils étaient difficiles à évaluer ou peu réalistes.

La grille a bien supporté cette première étape tout en suscitant chez plusieurs enseignantes du secteur l'intérêt et le désir de l'expérimenter dans leurs milieux respectifs, en salle d'urgence, en cardiologie, en chirurgie, en périnatalité.

Extraits de la grille d'observation à échelle descriptive pour le cours « Stage d'intégration » (180-620)

Critère : EXACTITUDE DES PLANS DE SOIN			
<i>COMPÉTENCE INSUFFISANTE</i>	<i>COMPÉTENCE MINIMALE</i>	<i>COMPÉTENCE MOYENNE</i>	<i>COMPÉTENCE SUPÉRIEURE</i>
<p>Les quatre derniers jours de stage, l'étudiante rédige la plupart des plans de soins au kardex en négligeant au moins deux des critères suivants :</p> <ul style="list-style-type: none"> • à partir de sa cueillette des données ; • en s'appuyant sur des principes scientifiques ; • en tenant compte des priorités. 	<p>Les quatre derniers jours de stage, l'étudiante rédige la plupart des plans de soins au kardex en respectant deux des critères suivants la plupart du temps :</p> <ul style="list-style-type: none"> • à partir de sa cueillette des données ; • en s'appuyant sur des principes scientifiques ; • en tenant compte des priorités. 	<p>Les quatre derniers jours de stage, l'étudiante rédige la plupart des plans de soins au kardex en respectant généralement les trois critères suivants :</p> <ul style="list-style-type: none"> • à partir de sa cueillette des données ; • en s'appuyant sur des principes scientifiques ; • en tenant compte des priorités. 	<p>Les quatre derniers jours de stage, l'étudiante rédige tous les plans de soins au kardex en respectant les trois critères suivants en tout temps :</p> <ul style="list-style-type: none"> • à partir de sa cueillette des données ; • en s'appuyant sur des principes scientifiques ; • en tenant compte des priorités.
Critère : PRÉCISION DANS LA RÉDACTION DU TRAVAIL DE RECHERCHE			
<i>COMPÉTENCE INSUFFISANTE</i>	<i>COMPÉTENCE MINIMALE</i>	<i>COMPÉTENCE MOYENNE</i>	<i>COMPÉTENCE SUPÉRIEURE</i>
<p>Au cours du stage, l'étudiante explique oralement à la professeure et aux pairs le travail de recherche pour au moins un de ses clients, mais le contenu contient des erreurs, est incomplet et/ou fait peu ou pas de liens entre ce qui est écrit et observé.</p>	<p>Au cours du stage, l'étudiante explique oralement à la professeure et aux pairs le travail de recherche pour au moins un de ses clients en abordant brièvement l'ensemble des rubriques prévues et en faisant occasionnellement des liens entre ce qui est écrit et observé.</p>	<p>Au cours du stage, l'étudiante explique oralement à la professeure et aux pairs le travail de recherche pour au moins un de ses clients en fournissant des détails sur la plupart des rubriques prévues et en faisant régulièrement des liens entre ce qui est écrit et observé.</p>	<p>Au cours du stage, l'étudiante explique oralement à la professeure et aux pairs le travail de recherche pour au moins un de ses clients en fournissant des détails sur toutes les rubriques prévues et en faisant toujours des liens entre ce qui est écrit et observé.</p>
Critère : DEGRÉ DE MOTIVATION			
<i>COMPÉTENCE INSUFFISANTE</i>	<i>COMPÉTENCE MINIMALE</i>	<i>COMPÉTENCE MOYENNE</i>	<i>COMPÉTENCE SUPÉRIEURE</i>
<p>Au cours du stage, l'étudiante refuse régulièrement de nouvelles expériences.</p>	<p>Au cours du stage, l'étudiante refuse occasionnellement de nouvelles expériences et s'en tient essentiellement à la tâche qui lui est assignée.</p>	<p>Au cours du stage, l'étudiante est, en tout temps, disponible pour de nouvelles expériences.</p>	<p>Au cours du stage, l'étudiante est, en tout temps, disponible pour de nouvelles expériences et démontre de l'initiative face à des expériences supplémentaires.</p>
Critère : PROMPTITUDE DANS LA GESTION DE L'ERREUR			
<i>COMPÉTENCE INSUFFISANTE</i>	<i>COMPÉTENCE MINIMALE</i>	<i>COMPÉTENCE MOYENNE</i>	<i>COMPÉTENCE SUPÉRIEURE</i>
<p>L'étudiante ne rapporte pas toujours une erreur dès qu'elle s'en rend compte, ou ne prend pas toujours les moyens pour la corriger.</p>	<p>L'étudiante rapporte, dès qu'elle s'en rend compte, une erreur et elle prend les moyens pour la corriger.</p>		

L'expérimentation proprement dite s'est déroulée durant le même trimestre, au dernier bloc de stage d'une durée de quatre semaines, à raison de quatre jours par semaine. Quatre enseignantes volontaires, sur une possibilité de huit, y ont participé.

Les enseignantes qui ont utilisé la grille on constaté que :

- les énoncés descriptifs, explicites et clairs, rendent l'observation plus facile en stage ;
- le jugement final est facilité par une collecte de données obligatoirement plus systématique en cours de stage ;
- les étudiantes font preuve de plus d'autonomie et d'initiative car elles connaissent les objets sur lesquels elles sont évaluées ;
- le jugement expert est ressenti comme plus objectif, tant par l'enseignante que par l'étudiante ;
- le jugement global reflète bien, de façon générale, le niveau de performance de chacune des étudiantes évaluées.

Les étudiantes, pour leur part, se sont dites plus satisfaites de l'usage de cette grille car la situation d'évaluation est rendue plus claire et plus transparente en situant les moments d'évaluation dans le temps et dans le contenu, en spécifiant de façon explicite les moments formatifs et sommatifs. Cela permet d'augmenter le lien de confiance de l'étudiante envers l'enseignante et favorise la relation centrée sur une vision commune des objectifs à atteindre. L'étudiante, sachant ce qu'il faut faire pour améliorer son niveau de compétence, développe son sens de l'autonomie et son initiative.

Suite à l'expérimentation, une version « finale » de notre grille d'observation à échelle descriptive, comportant 26 critères, a été adoptée par toutes les enseignantes concernées en mai 1997 en vue de la session d'automne 1997. Une réunion subséquente, à la fin de l'automne, nous a permis de continuer à améliorer la grille. À cette occasion, nous avons

convenu que l'évaluation formative en stage porterait sur les mêmes critères que l'évaluation sommative. Les étudiantes ont été informées de cette décision.

Conclusion

L'ensemble du travail de mise au point et d'expérimentation que nous décrivons ici a été fait avec la constante collaboration des membres du département de soins infirmiers de notre collège. Des copies du document de travail ont circulé dès la mi-session et une copie de la version finale a été remise à tous les membres du comité de travail en fin de session. Pendant cette même période, des ateliers de formation étaient tenus par le conseiller pédagogique sur les deux questions suivantes : comment tenir compte des variables affectives en évaluation sommative ? Comment s'assurer d'une évaluation sommative de qualité ?

On imaginera que ce chantier de travail a eu des répercussions à divers niveaux. Nous constatons aujourd'hui que le fait d'avoir élaboré la grille tout en informant et en consultant les membres du département et en organisant un atelier de formation sur le sujet a eu pour effet de les stimuler à amorcer ou à poursuivre leur réflexion au regard de l'évaluation, notamment en ce qui concerne : les éléments essentiels et fondamentaux à considérer lors de l'évaluation ; la préoccupation de s'assurer d'une certaine continuité dans les critères évalués d'un cours à l'autre ; la possibilité de s'entendre sur certains critères communs à l'ensemble du département ; la relation entre l'évaluation formative et l'évaluation sommative ; la pertinence des critères d'évaluation ; le choix des indicateurs de la compétence ; le calibrage entre ce qui décrit la compétence minimale et ce qui décrit la compétence insuffisante ; la pertinence d'adopter et de décrire des niveaux de compétence au-dessus du minimum ; etc.

Par ailleurs, d'autres équipes d'enseignantes ont entrepris d'expérimenter ce type de grille d'observation. Elles peu-

vent trouver du soutien auprès de nous, à cause de notre expérience, et le modèle que nous avons développé semble répondre à leurs besoins. De plus, ces mêmes équipes ont entrepris d'examiner, à partir de leurs critères de compétence, la progression des apprentissages proposés aux étudiantes d'un cours à l'autre. Cette réflexion départementale semble donc avoir stimulé un désir de cohérence sur certains aspects.

Au moment d'écrire ces lignes, nous sommes conscientes qu'il reste des problèmes à résoudre. Entre autres, il nous faut encore raffiner le libellé des énoncés descriptifs afin de trouver un équilibre entre une description trop ou pas assez détaillée. Il nous faut aussi examiner comment nos énoncés de compétence nous informent, après le stage, sur les lacunes et les points forts des étudiantes. Et, finalement, notre réflexion sur le calcul de la note porte sur les trois dimensions suivantes :

- La pondération des critères d'évaluation : les critères ont-ils vraiment tous la même importance relative ?
- La valeur accordée à chacun des niveaux de l'échelle : quel est le moyen le plus simple et le plus équitable pour calculer la note des stagiaires ? Comment distribuer les notes à chacun des niveaux de l'échelle de façon qu'on puisse conclure à la réussite ou à l'échec sans être confronté à une approche comptable ?
- Y a-t-il un critère pour lequel un échec entraînerait l'échec du stage dans son ensemble ? Combien faut-il réussir de critères, individuellement, pour que le stage soit réussi ?

Malgré les améliorations qu'il reste à y apporter, l'outil que nous avons mis au point s'avère déjà fort utile. Nous sommes convaincues que la grille d'observation à échelle descriptive convient particulièrement bien à l'évaluation des apprentissages en stages. Nous croyons en outre qu'elle pourrait servir à l'éva-

luation en laboratoire ou dans toute autre situation où il faut objectiver nos observations sur les étudiantes. ▣

RÉFÉRENCES

1. LEGENDRE, R., *Dictionnaire actuel de l'éducation*, Montréal, Guérin, 1993, p. 673.
2. Pour une étude détaillée sur les différents types d'échelles, voir : MORISSETTE, D., *Les examens de rendement scolaire*, Sainte-Foy, Presse de l'Université Laval, 1993, p. 201-202.
3. Il est fortement suggéré de ne pas dépasser cinq niveaux. Généralement, une échelle en quatre niveaux comportant un niveau sous le seuil de réussite (échec) suffit.
4. ORDRE DES INFIRMIÈRES ET INFIRMIERS DU QUÉBEC. *Les Perspectives de l'exercice de la profession infirmière*, Montréal, Mars 1996.