

Qualités psychométriques de l'entrevue orale structurée (EOS)

Par la Direction de l'amélioration de l'exercice

11 décembre 2015

La notion de validité ne réfère pas à une caractéristique intrinsèque d'un outil d'évaluation que l'on mesure une fois pour toutes¹, mais plutôt à un jugement évolutif que l'on pose en accumulant des éléments de preuve non seulement sur son contenu, mais également sur la manière dont il est utilisé et dans quel contexte, sur la justesse de l'interprétation des résultats et sur les conséquences de son utilisation. Dans son modèle, Messick propose cinq types d'éléments de preuve qui devraient être pris en compte lorsqu'il s'agit d'évaluer la validité d'un processus d'évaluation.

Éléments de preuve basés sur le contenu

Le processus de développement des cas à partir d'une table de spécifications contribue à sa validité en termes de contenu, dans la mesure où cette table est créée en s'assurant de bien représenter, d'une part, tous les domaines significatifs de la spécialité et, d'autre part, les éléments clés de la démarche clinique de cette spécialité.

Dans chaque spécialité, la banque de cas est non seulement revalidée régulièrement, mais également enrichie de nouveaux cas, ce qui permet de couvrir un plus large éventail de profils de pratique des candidats.

Éléments de preuve basés sur le processus de réponse

Bien que cet aspect n'ait pas été formellement testé, les membres des groupes d'experts considèrent que la résolution des cas fait appel aux mêmes processus cognitifs que ceux utilisés au quotidien par la majorité des médecins de leur spécialité.

De plus, la méthode de correction de l'EOS est basée, d'une part, sur le pourcentage de réponses attendues mentionnées par le candidat, et d'autre part, sur les commentaires émis par les évaluateurs sur chacun des cas, tels que l'aisance avec laquelle le candidat répond aux questions. Au fil des ans, des modifications ont été apportées à la méthode de correction pour en améliorer la justesse et la pertinence.

Éléments de preuve basés sur la structure interne

Ce type d'élément de preuve comprend, entre autres, la fidélité de l'outil, mesurée en termes de cohérence interne, de fidélité test-retest ou de fidélité interjuges, par exemple. Il concerne aussi la qualité des vignettes, soit leur pouvoir discriminant et leur degré de difficulté. À cet égard, des

¹ 1 Messick S. (1989). Validity. In: Linn R, editor, Educational Measurement (3rd ed.), New York: American Council on Education and Macmillan, p. 13-14.

analyses d'items ont été réalisées dans la plupart des spécialités et ont permis de cibler la révision de certaines vignettes.

C'est en médecine de famille que ce type d'élément de preuve a été le plus analysé, l'EOS dans cette spécialité étant la première à avoir été élaborée au Collège, au début des années 1990. Ainsi, la concordance interjuges de l'EOS en médecine de famille a été évaluée à 91,2 %, lors d'un projet pilote où six médecins sélectionnés de manière aléatoire ont été soumis chacun à 40 vignettes, en présence de deux évaluateurs. Plus récemment, le coefficient alpha de Cronbach, qui mesure la cohérence interne de l'EOS, a été évalué à 0,93 à partir d'un échantillon de 103 médecins de famille et de 46 vignettes.

La structure interne de l'EOS se rapporte également à sa marge d'erreur et à sa capacité à évaluer dans quelle mesure les résultats obtenus par un candidat peuvent être généralisés à l'ensemble des situations cliniques et des contextes auxquels il est confronté dans sa pratique. C'est au moyen d'une analyse de généralisabilité que l'on peut mesurer cet aspect de la validité. Pour l'instant, l'EOS n'a été évaluée qu'en médecine de famille, lors d'une étude réalisée sur un échantillon de 39 médecins et de 10 vignettes. L'étude a permis d'évaluer son coefficient de généralisabilité (coefficient G) à 0,94, avec une marge d'erreur de 10 % sur la note globale de l'EOS. Le coefficient de généralisabilité représente la proportion de variances « vraies » parmi les variances observées dans les résultats, soit la proportion de variances attribuables à de réelles différences entre les médecins évalués. La généralisabilité d'un outil d'évaluation est considérée comme satisfaisante lorsque ce coefficient est supérieur à 0,80.

Éléments de preuve basés sur la comparaison avec d'autres critères

Selon Messick, la validité d'un outil d'évaluation doit également être examinée en comparant ses résultats avec d'autres mesures du même phénomène. Dans le cas de l'EOS, censée mesurer la compétence d'un candidat, à défaut de disposer d'une mesure « étalon » de la compétence, on peut comparer les résultats de l'EOS avec l'opinion des experts qui ont évalué la compétence de ce candidat dans d'autres contextes. À cet égard, la concordance observée entre, d'une part, les résultats des candidats à l'EOS et, d'autre part, les évaluations faites par les pairs qui les supervisent dans leurs activités de remédiation après l'EOS, est rassurante et confirme la validité de l'EOS.

Éléments de preuve basés sur les conséquences de l'évaluation

Dans son modèle d'évaluation de la validité, Messick suggère de s'attarder aux conséquences de l'évaluation pour chacune des parties prenantes, en particulier pour le candidat. Cet aspect souvent négligé de la validité revêt une grande importance lorsque les enjeux liés à l'évaluation sont importants, comme c'est le cas pour l'EOS dont les conséquences peuvent être coûteuses pour un candidat qui se voit imposer, par exemple, un stage de perfectionnement de longue durée pendant lequel il devra cesser sa pratique (stage avec limitation d'exercice).

C'est donc un aspect important à évaluer et pour ce faire, on doit s'interroger sur les bénéfices attendus par rapport aux risques et inconvénients de l'évaluation. On doit également s'assurer que les décisions prises sur la base des résultats de l'EOS sont équitables, non biaisées, et qu'elles tiennent compte des valeurs de la communauté médicale et des exigences pertinentes sur le plan juridique.