Qualité psychométrique des QCM du concours de résidanat à la Faculté de médecine de Batna : Un outil d'amélioration des pratiques docimologiques

Psychometric quality of multiple-choice questions in the residency entrance examination at the Faculty of Medicine in Batna: A tool for improving assessment practices

Mohamed Ridha Guedjati[1], Hanane Benaldjia[1], Jalaleddinne Omar Bouhidel[1], Adeila Dallel Taibi[1], Abdesselam Hamizi[2]

1. Faculté de Médecine. Université Batna 2. Cité Ezzoouhour, Batna 05000 Algérie.
2. Faculté de Médecine. Université Constantine 3. Ali Mendjeli (La nouvelle ville). Constantine 25000. Algérie.

Résumé

Introduction: L'analyse psychométrique des QCM est une étape essentielle dans le processus d'évaluation. Elle permet de vérifier la validité et la fiabilité des items des QCM. Le concours de résidanat constitue un passage clé dans la formation médicale, visant à sélectionner les futurs médecins capables de répondre aux exigences cliniques et scientifiques. Notre objectif est d'identifier les caractéristiques psychométriques des QCM du concours de résidanat de médecine.
Méthode: 150 QCM ont été analysés à l'issue des épreuves du concours de résidanat de médecine session d'octobre 2025. Les épreuves ont inclus 410 candidats. Cette analyse a porté sur l'indice de difficulté (DIF I), l'indice de discrimination (DI) et le coefficient de l'alpha Cronbach.
Résultats: 78 (52 %) des QCM, présentaient un niveau de difficulté recommandé/acceptable, avec un DIF I moyen de 43± 2,03. 45 (30 %) étaient trop difficiles. 97 items (64 %) présentaient un DI entre excellent à acceptable et 40 (26 %) sont à rejeter, avec un DI moyen global de 0,325± 0,017. Une corrélation modérée et statistiquement significative a été observée entre le DIF I et le DI (r=0,36, p<0,01). La fiabilité des items, alpha de Cronbach, était satisfaisante (0,84).
Conclusion: Notre étude nous a permis d'identifier les QCM de qualité pouvant être intégrés à la banque de questions pour les évaluations futures, ainsi que les QCM nécessitant une révision. Cette rigueur dans l'analyse garantit la validité et la fiabilité des évaluations. Elle garantit une sélection rigoureuse conforme aux exigences académiques et pratiques.

Mots-clés : QCM, concours de résidanat, Psychométrie indice de difficulté, indice de discrimination

Abstract

Introduction: A psychometric analysis of multiple-choice questions (MCQs) is an important step in the assessment process. It makes sure that MCQ items are valid and reliable. The residency competition is an important part of medical training. It is how we choose future doctors who can meet the clinical and scientific requirements. Our aim is to find out what the medical residency competition's MCQs are like.
Methods: 150 multiple-choice questions (MCQs) were analysed from the October 2025 medical residency exam. There were 410 candidates in the exams. This analysis looked at the difficulty index (DIF I), the discrimination index (DI), and Cronbach's alpha coefficient.
Results: 78 (52%) of the multiple-choice questions were at a recommended/acceptable level of difficulty, with an average DIF I of 43± 2,03. 45 (30%) were too difficult. 97 items (64%) had a DI between excellent and acceptable, and 40 (26%) were rejected, with an overall average DI of 0,325± 0,017. There was a relationship between DIF I and DI (r=0.36, p<0.01). The reliability of the items was satisfactory at 0.84, as measured by Cronbach's alpha coefficient.
Conclusion: Our study allowed us to find good MCQs to add to the question bank for future tests, as well as MCQs that needed to be changed. This thorough analysis makes sure that the assessments are valid and reliable. It makes sure that the selection process is tough and that it meets the necessary academic and practical standards.

Keywords: MCQ, residency exam, psychometrics, difficulty index, discrimination index

Correspondance
Mohamed Ridha Guedjati
Faculté de Médecine. Université Batna 2. Cité Ezzoouhour, Batna 05000 Algérie.
Email: m.guedjati@univ-batna2.dz

LA TUNISIE MEDICALE-2026; Vol 104 (05): 627-633 DOI: 10.62438/tunismed.v104i05.6612

Cet article est distribué selon les termes de la licence internationale Creative commons Attribution-Pas d'Utilisation Commerciale 4.0 International(CC BY-NC-ND 4.0) qui autorise l'utilisation non commerciale, la production, la reproduction et la distribution, à condition que l'auteur original et la source soient mentionnés.

Introduction

L'évaluation joue un rôle important pour interpréter l'étendue des aptitudes des étudiants et leur progrès d'apprentissage (1). Le concours de résidanat constitue une étape cruciale dans le cursus médical, visant à sélectionner rigoureusement les futurs meilleurs médecins parmi un grand nombre de candidats. Les questions à choix multiples (QCM) sont couramment utilisées dans les évaluations des connaissances des étudiants en médecine. La qualité des QCM est importante en raison de son impact sur le niveau de compétence global des étudiants lors de leur évaluation. Assurer une évaluation juste et objective est fondamental pour garantir que seuls les candidats les plus compétents accèdent aux postes de spécialisation médicale, garant de la qualité des soins spécialisés. Des QCM bien conçues permettent d'évaluer des compétences cognitives de haut niveau, telles que l'interprétation, la pensée analytique et critique, l'application et la synthèse, conformément à la taxonomie de Bloom (2) et à la pyramide de Miller (3). L'élaboration de QCM est un défi pour les enseignants. Une banque de QCM soigneusement constituée après une analyse approfondie des items constitue un outil précieux afin d'établir une conformité docimologique des évaluations. Dans ce contexte, il est impératif que les QCM du concours de résidanat soient psychométriquement solides afin d'assurer la sélectivité et la fiabilité de ce processus de recrutement.

L'analyse des items d'un outil d'évaluation renseigne sur la validité et la fidélité de chaque item (4). L'analyse des items d'un QCM comprend l'indice de difficulté (DIF I) qui correspond au pourcentage d'étudiants ayant répondu correctement à l'item. L'indice de discrimination (DI) qui correspond à la capacité de distinguer les étudiants performants de ceux moins performants (4). Les QCM sont introduits au concours de résidanat en médecine en Algérie en 2012 dans une perspective visant à garantir une évaluation objective grâce à la correction optique automatisée. En plus, des valeurs métrologiques sont susceptibles d'être véhiculées par les QCM, telles que la validité, la fidélité et la fiabilité (5). Ces qualités sont indispensables pour garantir que la sélection repose sur des critères d'excellence stricts, renforçant ainsi la confiance dans le système de formation médicale spécialisée.

Objectif. Identifier les qualités psychométriques des QCM des épreuves du concours de résidanat de médecine session octobre 2025 à la faculté de médecine de Batna Algérie.

Methodes

Organisation générale du concours de résidanat

Le concours de résidanat comprend trois épreuves, de médecine, de chirurgie et de biologie. Ces épreuves couvrent les connaissances acquises au cours des six années de la phase pré-doctorale. Pour des raisons sanitaires de la pandémie COVID 19, le nombre des QCM a été réduit de 330 à 150 dont 50 pour chaque épreuve. La durée des épreuves est de 4 heures. Chaque QCM était composé d'un énoncé et de quatre options de réponse, dont une ou plusieurs sont correctes. La vérification de la validité externe des QCM est assurée par un jury d'examen désigné par le doyen parmi les enseignants de rang magistral. Les épreuves sont corrigées par un système de reconnaissance optique binaire. Chaque réponse correcte est notée sur 0,13. Les candidats sont classés par ordre de mérite.

L'analyse des items

Les 27 % des candidats ayant obtenu les meilleurs résultats ont été considérés comme forts (F), et les 27 % comme faibles (f) ayant obtenu les moins bons résultats. La présente étude porte sur l'ensemble des 410 candidats ayant réellement participé au concours, sur un total de 760 candidats inscrits. Chaque item a été analysé par les indices, de difficulté (DIF I) et de discrimination (DI). Pour l'étude de la fiabilité, nous avons utilisé le test alpha de Cronbach.

L'indice de difficulté (DIF I)

L'indice de difficulté DIF I représente le pourcentage des candidats, parmi le groupe des forts (F) et celui des faibles (f), ayant répondu correctement à la question. Ce pourcentage varie de 0 % à 100 %. Les critères de classification du DIF I sont les suivants : DIF I ≤ 20 % (trop difficile), DIF I entre 21 et 39 % (Difficile), DIF I entre 40 et 59% (Plage de difficulté idéale), DIF I entre 60 et 80 % (modérément facile) et DIF I > 80 très facile (6).

Le DIF I est calculé selon la formule, DIF I=(F+f/N)x100

L'indice de discrimination (DI)

L'indice de discrimination mesure la capacité d'un item à différencier les candidats ayant des aptitudes plus ou moins élevées et varie de 0 à 1. Cet indice permet de discriminer, les forts (F) des faibles (f) pour chaque QCM. Il a été calculé à l'aide de la formule DI = 2 × (F–f)/N, où F, f et N représentent les mêmes valeurs que celles mentionnées précédemment. Les critères de classification de l'DI sont les suivants : négatif (La question est trompeuse ou erronée, Besoin d'une révision ou d'un retrait), DI = 0 - 0,10 (à rejeter), DI = 0,10 – 0,19 (discrimination faible à améliorer), DI=0,20-0,29 (discrimination modérée à réviser), DI= 0,30-0,39 (discrimination acceptable) DI ≥ 0,40 (excellente discrimination) (6).

La fiabilité de la cohérence interne ou score de l'alpha de Cronbach

L'alpha de Cronbach est une mesure de la cohérence interne d'un test (6). La cohérence interne entre les trois épreuves (Biologie, Médecine et Chirurgie) a été évaluée au moyen du coefficient alpha de Cronbach, calculé sur la matrice des scores inter-épreuves. Il s'agit d'une mesure de la fidélité. Une valeur d'alpha comprise entre 0,8 et 0,9 est considérée comme satisfaisante. La fidélité désigne la capacité du test à mesurer ce qu'il est censé mesurer. Plus le score est proche de 1, plus le test est fidèle.

Analyse statistique

La comparaison des variables qualitatives a été faite à l'aide du test chi 2 de Pearson avec un risque d'erreur α, fixé à 5 %. La corrélation entre la note du cursus et la note du concours a été étudiée par le coefficient de corrélation de Pearson. L'analyse statistique a été réalisée à l'aide de MS Excel 2024 et SPSS, version 20.0 (IBM, Armonk, NY, États- Unis d'Amérique)

Résultats

L'analyse psychométrique de l'indice de difficulté des 150 QCM révèle une distribution globalement déséquilibrée des items. 48,7 % des QCM sont classées « difficiles » ou « trop difficiles », en faveur d'une sur-difficulté. 20 % des items se situent dans la zone de difficulté idéale nettement inférieure aux standards psychométriques recommandant (30 à 70 %). L'épreuve de biologie montre la calibration la moins adéquate, 52 % des items sont difficiles et 10 % dans la zone idéale. Les épreuves de médecine et de chirurgie présentent des profils légèrement plus équilibrés marqués par une proportion élevée d'items difficiles (48 % et 46 % respectivement) cette différence entre les 3 épreuves était statistiquement non significative (p=0,432) (Tableau 1).

Tableau 1. Répartition des QCM selon l'indice de difficulté aux trois épreuves du concours de résidanat.

L'épreuve	Trop difficile	Difficile	Difficulté idéale	Modérément facile	Très facile	Total	p
Médecine	12 (24,0 %)	12 (24,0 %)	14 (28,0 %)	10 (20,0 %)	2 (4,0 %)	50 (100 %)	0,432
Chirurgie	11 (22,0 %)	12 (24,0 %)	11 (22,0 %)	10 (20,0 %)	6 (12,0 %)	50 (100 %)	
Biologie	10 (20,0 %)	16 (32,0 %)	5 (10,0 %)	13 (26,0 %)	6 (12,0 %)	50 (100 %)	
Total	33 (22,0 %)	40 (26,7 %)	30 (20,0 %)	33 (22,0 %)	14 (9,3 %)	150 (100 %)	

L'analyse de l'indice de discrimination des 150 QCM montre une distribution globalement satisfaisante marquée par des limites psychométriques. 40,7 % des items présentent une discrimination excellente. 4,7 % des items sont classés comme QCM trompeuses et 12 % présentent une mauvaise discrimination, ce qui représente près de 17 % d'items dysfonctionnels. 17,3 % atteignent un niveau acceptable de discrimination. Les comparaisons entre épreuves révèlent que 46 % de l'épreuve de biologie sont des items excellents suivie de la chirurgie (42 %) et de la médecine (34 %) (Tableau 2).

Tableau 2. Répartition des QCM selon l'indice de discrimination aux trois épreuves du concours de résidanat

L'épreuve	Question trompeuse	Mauvaise discrimination	Discrimination faible	Discrimination modérée	Discrimination acceptable	Excellente discrimination	Total	p
Médecine	1 (2 %)	6 (12 %)	8 (16 %)	5 (10 %)	13 (26 %)	17 (34 %)	50 (100 %)	0,343
Chirurgie	1 (2 %)	6 (12 %)	8 (16 %)	6 (12 %)	8 (16 %)	21 (42 %)	50 (10 %)	
Biologie	5 (10 %)	6 (12 %)	4 (8 %)	7 (14 %)	5 (10 %)	23 (46 %)	50 (100 %)	
Total	7 (4,7 %)	18 (12 %)	20 (13,3 %)	18 (12 %)	26 (17,3 %)	61 (40,7 %)	150 (100 %)	

L'analyse croisée des indices de difficulté et de discrimination des 150 QCM montre une relation cohérente entre l'indice de difficulté des items et l'indice de discrimination. Les items trop difficiles présentent une faible discrimination. 57,6 % trompeurs ou de mauvaise discrimination. Aucun item ne présente une discrimination excellente. Les items très faciles sont peu discriminants. 92,9 % des items ont une discrimination faible ou modérée. Les items de difficulté idéale et difficile affichent les meilleures performances psychométriques, respectivement 63,3 % et 62,5 % des questions à discrimination excellente, et les items modérément faciles restent globalement discriminants (51,5 % excellente, 39,4 % acceptable) et cette différence était statistiquement significative (p<0,0001) (Tableau 3).

Une bonne cohérence interne entre les épreuves du concours a été constatée. L'Alpha Cronbach à 0,84 suggère que les notes obtenues aux différentes épreuves sont globalement homogènes et mesurent une même compétence sous-jacente (Tableau 4).

Le fait que l'alpha standardisé est plus élevé (0,897) confirme une cohérence est très forte. Les corrélations entre les différentes épreuves et la note du concours vont de 0,49 à 0,92, toutes sont modérées à très fortes.
L'analyse de l'éventuelle corrélation entre la note obtenue au concours avec celle obtenue à la fin du cursus, a montré que 21,9 % (R^2=0,219) de la variance des notes du cursus est corrélée à la note du concours, r = 0,468 ; p < 0,0001 (Figure 1).

Tableau 3. Relation entre la classe de l'indice de difficulté et la classe de l'indice de discrimination des QCM du concours de résidanat

Classe de l'indice de difficulté	Classe de l'indice de discrimination						Total	p
	Question trompeuse	Mauvaise discrimination	Discrimination faible	Discrimination modérée	Discrimination acceptable	Excellente discrimination		
Trop difficile	7 (21,2 %)	12 (36,4 %)	7 (21,2 %)	5 (15,2 %)	2 (6,1 %)	0	33 (100 %)	< 0,0001
Difficile	0	4 (10 %)	2 (5 %)	4 (10 %)	5 (12,5 %)	25 (62,5 %)	40 (100 %)	
Difficulté idéale	0	1 (3,3 %)	3 (10 %)	2 (6,7 %)	5 (16,7 %)	19 (63,3 %)	30 (100 %)	
Modérément facile	0	1 (3 %)	1 (3 %)	1 (3 %)	13 (39,4 %)	17 (51,5 %)	33 (100 %)	
Très facile	0	0	7 (50 %)	6 (42,9 %)	1 (7,1 %)	0	14 (100 %)	
Total	7 (4,7 %)	18 (12 %)	20 (13,3 %)	18 (12 %)	26 (17,3 %)	61 (40,7 %)	150 (100 %)	

Tableau 4. Évaluation de la cohérence interne

Variables	m±ET	r Médecine	r Chirurgie	r Biologie	r Note Concours	Alpha de Cronbach	Alpha de Cronbach standardisé
Médecine	8,11±2,47	1	0,494	0,717	0,889	0,84	0,89
Chirurgie	9,20±4,62	0,494	1	0,496	0,589		
Biologie	8,70±2,70	0,717	0,496	1	0,922		
Note Concours	8,62±2,43	0,889	0,589	0,922	1		

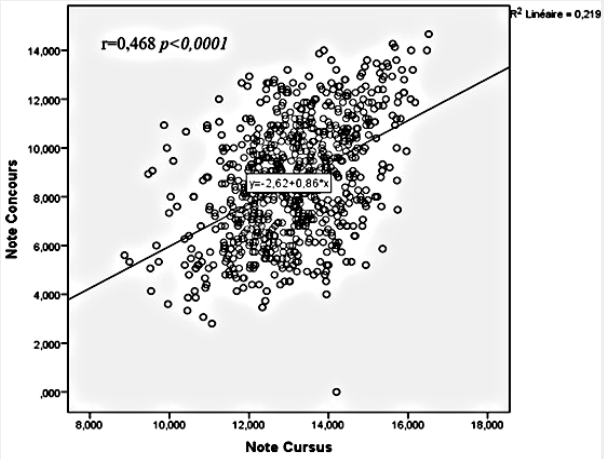


Figure 1. Corrélation entre la note du cursus et la note du concours.

Discussion

Le processus de sélection pour le programme de résidence en médecine peut prendre plusieurs formes. Il représente une opportunité pour les candidats d'accéder à un poste de formation spécialisée (5,7). Ce concours joue un rôle déterminant dans la formation des futurs médecins spécialistes, garants de la qualité des soins médicaux à venir. Dans le cadre du processus de sélection, les candidats qui satisfont aux critères d'éligibilité sont classés par ordre de mérite. Les lauréats de cette procédure ont l'opportunité de poursuivre une formation spécialisée de quatre à cinq ans, leur permettant d'exercer en tant que médecins spécialistes. Le concours marque un tournant stratégique dans le choix de carrière des candidats (5). Par conséquent, garantir l'excellence et l'équité de ce concours est fondamental pour identifier les talents médicaux les plus prometteurs. Le processus de sélection pour le programme de résidence médicale se décline en trois étapes épreuves, de médecine, de chirurgie et de biologie. Le processus d'évaluation s'effectue à l'aide d'examens utilisant des QCM. Dans leur préparation aux épreuves du concours de résidanat de médecine, les candidats se réfèrent au programme spécifique qui leur est destiné. Le processus d'évaluation des épreuves est assuré par un jury spécialisé qui se consacre à la conception, à la gestion et à la validation des sujets d'examen. Le rôle de ce jury est de garantir la conformité externe des QCM par rapport au programme de préparation au concours, ce qui est essentiel pour maintenir l'intégrité du processus d'évaluation. Cette rigueur garantit, en une part, que le concours reste un instrument juste et fiable pour la sélection des futurs professionnels de santé. Cependant, il est très important d'observer les critères psychométriques de l'évaluation, la validité, la fiabilité (8). La validité exprime la qualité externe d'un test, souvent dans une visée prédictive, par la corrélation de la mesure obtenue avec une autre mesure objective jugée pertinente (8, 9). Dans le contexte du concours, la validité garantit que les résultats traduisent fidèlement les compétences réelles des candidats. La fidélité repose sur la structure interne du test (10). Ces deux vertus docimologiques prennent appui dans un domaine complexe de la psychologie. Assurer la fiabilité des épreuves est donc essentiel pour que les décisions liées au classement et à la sélection soient équitables et reproductibles. L'analyse des items est une procédure relativement simple et précieuse qui permet d'analyser les observations, d'interpréter les connaissances acquises par les candidats et d'obtenir des informations sur la qualité des questions d'examen

(11, 12). Elle est indispensable dans un concours aussi stratégique que le résidanat, pour affiner la qualité des questions et optimiser la sélection des meilleurs profils.

Le concours de résidanat de médecine utilise le type de QCM où le candidat est censé démêler les réponses justes de celles fausses (13). L'efficacité de l'évaluation par QCM repose entièrement sur la qualité des questions d'examen. Nous avons jugé utile d'analyser les QCM des épreuves du concours de résidanat de médecine en post - examen. Une telle analyse est plus qu'une obligation. La qualité des questions est directement corrélée à la pertinence du classement final, ce qui impacte la formation et le futur exercice médical. L'identification des caractéristiques psychométriques de validité interne et de fiabilité des épreuves, de manière objective, renforce les qualités du processus d'évaluation (14).

Sur 150 QCM, 48,7 % sont difficiles ou trop difficiles, témoignant d'une sur-difficulté susceptible de réduire la sensibilité de l'épreuve et d'affecter ses capacités discriminantes. Cette situation constitue un écueil pour un processus d'une importance aussi déterminante que le recrutement des futurs médecins spécialistes. 20 % des items se situent dans la zone de difficulté idéale, proportion nettement inférieure aux standards psychométriques recommandant que 30 à 70 % des items présentent une difficulté optimale afin d'assurer une mesure valide et fiable des performances. Cette configuration suggère la nécessité d'une révision systématique de la banque des QCM afin d'améliorer la qualité métrique de l'évaluation.

Un nombre de travaux, portant sur des études similaires, rapporte des valeurs d'indice de difficulté au- delà de 80 %. L'étude de Pande rapporte que 85 % des QCM présentaient un niveau acceptable de DIF I (30–70 %) (12). De nombreuse études ont montré que 80 % des QCM avaient un niveau acceptable de DIF I (15). Date et al a montré que 70 % des questions avaient un niveau acceptable de DIF I, 20 % étaient trop faciles et 10 % trop difficiles (11). Cette capacité discriminante est essentielle pour un concours comme le résidanat, qui doit identifier avec une certaine précision les meilleurs candidats destinés à devenir médecins spécialistes. Dans notre étude 87 (58 %) des items présentent une discrimination acceptable à excellente ce qui constitue une indication acceptable et adéquate de différencier les candidats les plus performants des moins performants. Près de 17 % d'items présentent un dysfonctionnement susceptible d'altérer la validité de l'évaluation. De même 17,3 % atteignent un niveau de discrimination jugé acceptable. Ces résultats soulignent la nécessité d'une révision ciblée des items à faible ou mauvaise discrimination, ainsi que des questions trompeuses, afin d'optimiser la qualité psychométrique de l'épreuve et d'améliorer sa capacité à mesurer de manière fiable et discriminante les compétences des candidats (11, 16). Date et d'autres auteurs ont rapporté des résultats proches. 78 % des items présentaient un pouvoir discriminant acceptable à excellent (DI > 0,20), 65 à 70 % un pouvoir discriminant bon à excellent (DI ≥ 0,25) et 24 % un faible pouvoir discriminant (DI ≤ 0,20) (11, 12).

L'interprétation des DI des items doit tenir compte du type de test analysé. Les questions dont l'DI est faible sont souvent formulées de manière ambiguë et devraient être réexaminées. Une valeur d'ID plus élevée indique une plus grande efficacité d'une question à choix multiple (17).

La présente étude montre que les items trop difficiles présentent une faible discrimination, indiquant qu'une augmentation du DIF I diminue la capacité à discriminer les candidats ayant obtenu des scores élevés et faibles, et que le nombre de questions faciles utilisées dans les tests QCM est réduit. Plusieurs études ont également mis en évidence une faible relation entre le DIF I et le DI (18, 19). La recherche bibliographique ramène des variabilités concernant les indices de difficulté et de discrimination étudiés. Ceci peut être expliqué en une part par l'hétérogénéité des domaines étudiés. Certaines études portent sur une analyse des items de QCM d'une discipline (20-22), d'autres s'intéressent à des items des épreuves classantes ou des concours (5, 9, 23), comme le cas de notre travail. Il est fort possible aussi que le niveau cognitif évalué a un rôle dans cette variabilité (24).

Si les QCM restent des outils d'évaluation valides, objectives et fiables, leur utilisation dans les épreuves classantes comme le concours de résidanat de médecine, soulève quelques écueils. Les niveaux cognitifs évalués (4). Cet aspect n'a pas été traité dans ce travail. Il serait plus intéressant d'évaluer, si les différents niveaux taxonomiques de Bloom, ont été examinés par les QCM. Dans une réflexion publiée en 2024 (13), il a été proposé d'intégrer, outre les QCM du type A qui supportent la meilleure réponse (25, 26), des QCM évaluant les niveaux taxonomiques plus élevés. Ces QCM peuvent être à contexte clinique ou faisant partie des Key Feature Problem KFP (27). Les épreuves classantes se basent sur l'hypothèse que les meilleurs éléments qui ont l'habitude de montrer une performance durant leur cursus, sont les mêmes qui obtiennent les meilleures notes à ces épreuves. Ce qui valide la cohérence entre la formation et l'évaluation finale. 22 % des candidats performant durant le cursus le sont aussi au concours de résidanat. Il est fort possible que le concours évalue aussi d'autres compétences (par exemple la préparation spécifique, la gestion du stress, la rapidité de raisonnement, etc.) (28).

Nous avons été confrontés à un certain nombre de limites. La base de données des épreuves était insuffisante. On n'était pas en mesure de calculer le Kuder-Richardson 20 (KR-20) (29) ni de coefficient alpha de Cronbach par item. Ces paramètres paraissent plus pertinents que le coefficient alpha de Cronbach par épreuve dont il dérive (30, 31). Cette limite ne semble pas affecter la mesure de la cohérence interne des épreuves. Ces dernières semblent présenter une cohérence entre modérée et satisfaisante. De manière générale, la fiabilité de la cohérence interne de notre travail est pratiquement concordante avec celle de plusieurs études (32). Desy dans son étude portant sur l'évaluation de la fiabilité de la cohérence interne des épreuves 2016-2021, avait constaté une variabilité de l'alpha Cronbach de 0,50 à 0,70 (33). De même des résultats assez proches ont été observés dans d'autres travaux (34). Notre étude a touché 150 QCM ce qui représente 5 % de la banque depuis l'adoption de la

correction optique automatisée en 2012. Mustafa, dans son étude a analysé 155 QCM d'une banque de 3 années entre 2020 et 2022 (35).

Conclusion

L'évaluation psychométrique des QCM constitue un outil clé pour garantir la validité et la fiabilité des épreuves du concours de résidanat. La rigueur dans l'analyse des items assure une sélection objective et précise des candidats, condition essentielle pour identifier les futurs meilleurs médecins. En mettant en lumière les forces et les limites des QCM, cette analyse permet d'améliorer continuellement la qualité des outils d'évaluation.

Références

1. Epstein RM. Assessment in medical education. New England journal of medicine. 2007;356(4):387-96.
2. Adams NE. Bloom's taxonomy of cognitive learning objectives. Journal of the Medical Library Association: JMLA. 2015;103(3):152.
3. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. International Journal of Medical Education. 2019;10:191.
4. Abey Abraham Joy, Neethu Soman, HymaV, Jiyo Chacko, Sujatha M.B., Sreelakshmi Venugopal. Item Analysis of Multiple-Choice Questions Using Difficulty and Discrimination Indices International Journal of Life Sciences, Biotechnology and Pharma Research;2025;14(9).
5. Mohammed OEAH, Srinivasamurthy SK, Bhat R, Alhassan Eltahir YAM, Hamdy Elshamly HA, Mohammed F, Hamad B. Postgraduate Clinical Residency: The Impact of Multiple-Choice Question Quality on Exam Success Rates. Advances in Medical Education and Practice. 2025:1381-97.
6. Vaske JJ, Beaman J, Sponarski CC. Rethinking internal consistency in Cronbach's alpha. Leisure sciences. 2017;39(2):163-73.
7. Eleragi A, Miskeen E, Hussein K, Rezigalla AA, Adam MI, Al-Faifi JA, et al. Evaluating the multiple-choice questions quality at the College of Medicine, University of Bisha, Saudi Arabia: a three-year experience. BMC Medical Education. 2025;25(1):233.
8. Mirbahai L, Adie J. Applying the utility index to review single best answer questions in medical education assessment. Archives of Epidemiology and Public Health. 2020;1:1-5.
9. Yahia AIO. Post-validation item analysis to assess the validity and reliability of multiple-choice questions at a medical college with an innovative curriculum. Natl Med J India. 2021;34(6):359-62.
10. Newton, P.E. (2013), Two Kinds of Argument? Journal of Educational Measurement, 50: 105-109.
11. 11. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. Int J Basic Clin Pharmacol. 2019;8(9):1999-2003.
12. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. South-East Asian Journal of Medical Education. 2013;7(1).
13. Guedjati MR, Bouhidel OJ, Benaldjia H, Derdous C. The Medical Residency Examination in the Era of the Overhaul of Medical Study Programs, a Call for Reform. Reflections for New Docimological Orientations. Acta Scientific Medical Sciences. 2024 May 1;63–8.
14. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Medical education. 2005;39(3):309-17.
15. Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. Research Journal of Pharmaceutical Biological and Chemical Sciences. 2015;6(3):1242-51.
16. National Board of Medical Examiners. NBME Item-Writing Guide: Constructing Written Test Questions for the Health Sciences. National Board of Medical Examiners; 2024.
17. Mitra NK, Chitra E, Wong PS, Sow CF, Binti Abd Razak SS, Er HM, Nadarajah VD. Artificial Intelligence: Opportunities and Challenges in Health Professions Education. International E-Journal of Science, Medicine & Education. 2023;17(3).
18. Kelley TL. The selection of upper and lower groups for the validation of test items. Journal of educational psychology. 1939;30(1):17.
19. Christian DS, Prajapati AC, Rana BM, Dave VR. Evaluation of multiple choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat. Int J Community Med Public Health. 2017;4(6):1876-81.
20. Zubairi NA, AlAhmadi TS, Ibrahim MH, Hegazi MA, Gadi FU. Effective use of Item Analysis to improve the Reliability and Validity of Undergraduate Medical Examinations: Evaluating the same exam over many years: A different approach. Pakistan Journal of Medical Sciences. 2025;41(3):810.
21. Gebremichael MW, Baraki B, Mehari M-A, Assalfew B. Item analysis of multiple choice questions from assessment of health sciences students, Tigray, Ethiopia. BMC Medical Education. 2025;25(1):441.
22. Wrzesińska MA, Rakoczy J, Binder-Olibrowska KW, Kostyła M, Allott VE, Harris SR, et al. Comparing structured note-taking and multiple-choice question writing as learning strategies among physiotherapy students. BMC Medical Education. 2025;25(1):1665.
23. Srisomsak V, Sitticharoon C, Keadkraichaiwat I, Meethes S, Inpaen I. Detection of flawed multiple-choice questions in preclinical medical education using item difficulty and discrimination indices: a six-year analysis. BMC Medical Education. 2025.
24. Hwang R, Paley M, Bull D, Omrani O, Vergara-Jalandoni D, Egbury G. A Systematic Review and Meta-Analysis on the Impact of Peer-Made MCQ Question-Bank Usage on Summative Assessments in Medical Education. Journal of Advances in Medical Education & Professionalism. 2025;13(4):259.
25. Rahim AFA, Simok AA, Wahab SFA. A Guide for Writing Single Best Answer Questions to Assess Higher-Order Thinking Skills Based on Learning Outcomes. Education in Medicine Journal. 2022;14(2).
26. Oc Y, Hassen H. Comparing the effectiveness of multiple-answer and single-answer multiple-choice questions in assessing student learning. Marketing Education Review. 2025;35(1):44-57.
27. Kwa SK, Tan CW, Ng AC, Mobarak SMAS, Selamat HA. Recommendations and tips for passing the key feature problem examination. Malaysian Family Physician: the Official Journal of the Academy of Family Physicians of Malaysia. 2022;17(2):2.
28. Guedjati MR. Le concours de résidanat de médecine, peut-il servir à autre chose qu'au classement des candidats ? Oral Communication Pédagogie Médicale 10 S25-S67 (2009)
29. Uyanah DA, Nsikhe U. The theoretical and empirical equivalence of Cronbach alpha and Kuder-Richardson formular-20 reliability coefficients. International Research Journal of Innovations in Engineering and Technology. 2023;7(5):17.
30. Cronbach LJ. Coefficient alpha and the internal structure of tests. psychometrika. 1951;16(3):297-334.
31. Edelsbrunner PA, Simonsmeier BA, Schneider M. The Cronbach's alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies. Educational Psychology Review. 2025;37(1):4.
32. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Medical Journal Armed Forces India. 2021;77:S85-S9.
33. Desy J, Harvey A, Weeks S, Busche KD, Martin K, Paget M, et al. The impact of systematically repairing multiple choice questions with low discrimination on assessment reliability: an interrupted time series analysis. Canadian Medical Education Journal. 2024;15(3):52-6.
34. Lertsakulbunlue S, Kantiwong A. Development and validation of immediate self-feedback very short answer questions for medical

students: practical implementation of generalizability theory to estimate reliability in formative examination designs. BMC Medical Education. 2024;24(1):572.

35. Mustafa, S., Hamid, O.E. Psychometric item/question analysis of multiple-choice questions in fixed prosthodontics exam. BMC Med Educ (2025).