

Fidélité, équité et impact des mini entretiens multiples dans le processus de sélection des étudiants en santé en France

Reliability, fairness and impact of multiple mini interviews in the selection of health students in France

Antoine JACQUET*, Jérémie RIOU, Elisabeth LETERTRE, Catherine PASSIRANI, et Jean-Paul SAINT-ANDRÉ

Université d'Angers, 40 rue de Rennes, BP 73532, 49035 Angers Cedex 01, France

Manuscrit soumis le 7 avril 2018 ; accepté pour publication le 13 novembre 2018

Résumé. Contexte : L'Université d'Angers (UA) a créé le parcours pédagogique PluriPASS en remplacement de la PACES (première année commune aux études de santé) et a introduit, pour la première fois, les mini entretiens multiples (MEM) dans le processus de sélection des étudiants. Le but de cette étude est d'évaluer la fidélité des MEM, d'en vérifier l'équité et d'en mesurer l'impact. **Sujets/matériel/méthodes :** 294 étudiants ont été convoqués aux MEM en juin 2016. Les MEM étaient composés de 4 stations évaluant 6 compétences. Les critères suivants : âge, sexe, redoublement, aide boursière, catégorie socio professionnelle des parents, mention au baccalauréat, moment de l'évaluation et scénario utilisé, ont été étudiés par analyse univariée puis multivariée. L'impact des MEM a été évalué par la méthode des rangs limites. **Résultats :** Les coefficients alpha de Cronbach pour les 6 compétences évaluées s'élevaient de 0,612 à 0,935. Le sexe masculin était associé à un meilleur score global aux MEM (+ 27,4 points IC95 % = [9,2 ; 45,7]). Les étudiants ayant obtenu une mention très bien au baccalauréat ont obtenu de meilleurs scores globaux (+ 30,41 points ; IC95 % = [14,75 ; 46,12]). Il n'y avait pas de différence sur les autres variables socio démographiques. Il n'a pas été retrouvé de différence en fonction des scénarios. Cinquante-trois étudiants différents ont été exclus du classement utile d'une filière suite aux MEM. **Conclusion :** Les MEM au sein de l'expérimentation PluriPASS sont cohérents et présentent un intérêt pédagogique et sélectif. Néanmoins une attention particulière est nécessaire pour en garantir l'équité, dans leur mise en œuvre comme dans les travaux de recherche futurs.

Mot clés : mini entretiens multiples, MEM PluriPASS, sélection en santé, éducation en santé, admission en santé, PACES

Abstract. Context: The University of Angers (UA) created the PluriPASS cursus that replaced the PACES (première année commune aux études de santé/First year of common health studies) and introduced multiple mini-interviews (MMIs) in the selection process of students in France. This study is designed to evaluate MMI reliability, fairness and impact. **Topics/materials/methods:** Some 294 students attended the MMIs in June 2016. The MMIs were made up of four stations evaluating six competences. The following criteria (age, gender, novice/repeater, scholarship, parents' socio-professional category, bachelor's degree with honors, time of the evaluation and scenario) were examined using both univariate and multivariate analysis. The impact of the MMIs was evaluated following the limited-ranges method. **Outcome:** Cronbach's alpha coefficients for the 6 evaluated competences ranged from 0.612 to 0.935. Males ranked better overall with MMIs (+ 27.4 pts; CI95% = [9,2; 45,7]). Students who were got their bachelor's degree with high honors ("Très Bien" distinction, higher than 16/20) also had a better overall score (+ 30.41 pts; CI95% = [14.75; 46.12]). There were no differences with respect to the other socio-demographic variables, and none regarding the different scenarios considered. Following the MMI follow-ups, 53 different students were excluded from the ranking in a study path. **Conclusion:** As part of the PluriPASS testing, MMIs were found to be coherent as well as educationally and selectively interesting. However, they require particular consideration to ensure fairness, both during implementation and in future research studies.

Keywords: multiple mini-interviews, MMI, PluriPASS, medical student admission, medical education, medical student selection

*Correspondance et offprints : Antoine JACQUET, Université d'Angers, 17 allée de la Palombière, 44980 Sainte Luce sur Loire, France. Mailto: antoinejacquetangers@gmail.com.

Introduction

L'Université d'Angers (UA) expérimente depuis septembre 2015 un nouvel accès aux études de santé, suite à un appel à projets lancé par le Ministère de l'Enseignement Supérieur et de la Recherche en 2013. Ce projet comporte des épreuves orales – les mini entretiens multiples (MEM) – pour l'admission dans les filières de santé, selon des modalités déjà expérimentées dans plusieurs pays mais jamais utilisées en France jusqu'alors. Nous décrivons dans cet article les modalités d'organisation de ces MEM et analysons les résultats observés, en termes de fidélité, d'équité et d'impact.

Contexte

L'expérimentation de l' Université d'Angers : le parcours « PluriPASS »

PluriPASS est un portail d'entrée pluridisciplinaire vers les études de santé contingentées mais aussi vers d'autres filières de l'enseignement supérieur. Il se structure en quatre semestres (S1 à S4) au cours desquels les étudiants participent à un processus d'orientation active. Ce processus s'inscrit dans un cycle de licence, lui-même inscrit dans la perspective « Licence-Master-Doctorat » (LMD), en lien avec les orientations européennes du processus de Bologne. Au cours des deux années du parcours PluriPASS, l'étudiant peut, en fin de semestre, s'orienter vers une filière LMD ou un cursus d'ingénieur de son choix. Les objectifs de PluriPASS sont de garantir un recrutement efficace dans six filières contingentées de santé (médecine, pharmacie, maïeutique, odontologie, kinésithérapie, ergothérapie) tout en offrant la possibilité à tout étudiant qui valide ses semestres de poursuivre des études supérieures. PluriPASS permet ainsi d'éviter le « gâchis humain » en reconnaissant les acquis de tous les étudiants et en favorisant la poursuite d'études supérieures longues. PluriPASS décourage le redoublement des étudiants puisqu'il n'est pas possible de se présenter au concours d'entrée dans les filières contingentées alors que l'on redouble son année.

Organisation des enseignements de « PluriPASS »

Les semestres sont constitués autour de trois blocs d'enseignements : sciences de la vie, sciences de l'ingénieur, sciences humaines et sociales, représentés sur la [figure 1](#). Lors de la première année (S1 et S2), les enseignements portent sur l'ensemble des trois blocs au travers de douze unités d'enseignement (UE). Une autre sélection à lieu à la fin du semestre 3. Enfin, lors du semestre 4, les étudiants qui ne sont pas entrés en filière de santé choisissent une filière, préfigurant le choix du parcours qu'ils intégreront lors du semestre 5 (c'est à dire en licence 3, en dehors du parcours PluriPASS). D'autres UE sont dispensées dans le parcours PluriPASS. Il s'agit d'une part, d'UE optionnelles (UEO) et, d'autre part, du projet personnel et professionnel de l'étudiant (3PE). Dans ce cadre, chaque étudiant doit élaborer deux projets au cours des semestres 1 et 2.

Processus de sélection

Les étudiants ayant validé (obtenu des notes suffisantes) les deux premiers semestres disposent d'une première chance d'intégrer les études de santé. Ceux qui n'ont pas validé et décident de redoubler ne pourront pas se présenter au concours après avoir validé le S2 mais seulement après avoir validé le S3. Soixante-quinze pour cent des effectifs déterminés par le *numerus clausus* sont recrutés à la fin du S2. À la fin du S2, 50 % des étudiants sont recrutés sur des critères académiques purs (admission directe à l'issue du concours écrit), et 50 % sur des critères académiques et sur leurs performances aux mini entretiens multiples, en classant les étudiants selon la méthode du rang moyen ($\text{rang moyen} = \frac{\text{rang écrit} + \text{rang MEM}}{2}$).

À la fin du S3, les 25 % restants des effectifs prévus par le *numerus clausus* sont recrutés, après une nouvelle série de MEM.

Les mini entretiens multiples (MEM)

Les MEM ont été créés en 2002 à l' Université de McMaster [1]. De multiples articles ont montré leur cohérence (validité interne), leur reproductibilité et leur validité externe. Leur faisabilité, ainsi que leur acceptabilité et leur équité ont également été reconnues dans le monde Anglo-Saxon [2,3,4].

L'épreuve se compose de stations (ou étapes) au cours desquelles l'étudiant va préparer un entretien, être évalué (entretiens) ou va se reposer. Le nombre de stations, les modalités d'évaluation, de préparation, l'organisation des parcours sont variables d'une épreuve à l'autre. De nombreuses modalités d'entretiens ont été testées, par exemple avec un ou plusieurs examinateurs, ou avec des acteurs pour des mises en situation. L'évaluation de chaque entretien est structurée autour d'une grille d'évaluation standardisée.

Les MEM entrent donc en rupture avec la sélection traditionnelle française en médecine : ils ne sont pas écrits et n'évaluent pas les résultats des seules capacités de mémorisation.

Objectifs du travail

Trois paramètres semblaient pertinents à évaluer prioritairement : la fidélité, pour vérifier la cohérence interne des épreuves dans leur objectif de mesurer précisément une compétence unique ; l'équité, pour vérifier que l'épreuve n'est pas biaisée par des éléments extérieurs aux seules compétences des étudiants ; l'impact, pour analyser les conséquences pratiques des MEM sur les classements dans un tel processus. D'autres éléments semblaient également importants à analyser : la validité des épreuves, c'est à dire leur pertinence à prédire d'autres résultats, l'impact des MEM sur le travail des étudiants ainsi que sur leur état d'esprit, la valeur prédictive des MEM sur la suite des études et une étude de coût.

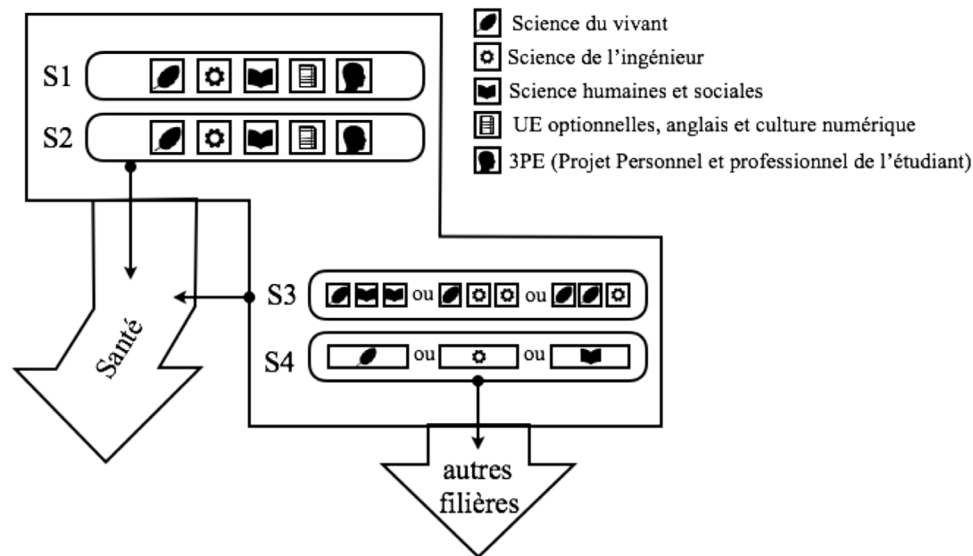


Figure 1. Organisation de PluriPASS. UE : Unité d'enseignement ; S : Semestre.

Méthodes

Cette étude porte sur les MEM du deuxième semestre de la première année « PluriPASS » qui se sont déroulés en juin 2016.

Population

La population de PluriPASS de référence était celle inscrite au mois de janvier en deuxième semestre de PluriPASS.

Les étudiants de PluriPASS étaient majoritairement des étudiants post-baccalauréat (le baccalauréat étant le diplôme de fin d'études secondaires en France), âgés de 18 à 19 ans, titulaires d'un baccalauréat scientifique (série S). Sur 1161 étudiants inscrits au S2 de PluriPASS, 294 ont été convoqués aux MEM. Six absents ont été exclus de l'analyse ; 72 % étaient des femmes (contre 71 % dans la population de PluriPASS) ; 31 % étaient boursiers (contre 42 %) ; 44 % étaient primants (contre 59 %) ; enfin 29 % avaient reçu une mention « Très Bien » ou mieux au baccalauréat contre 19 % dans la population de PluriPASS. La [figure 2](#) représente le processus de l'étude.

Les modalités de l'épreuve

L'épreuve comprenait quatre stations de 10 minutes chacune (cinq minutes de présentation et cinq minutes de discussion ou sept + trois minutes) dont la note pouvait aller de 6 à 100. La construction du score était réalisée par l'addition de sous-scores pour 5 + 1 items. Les cinq premiers items correspondaient à des compétences évaluées par une échelle à cinq grades. Les points attribués en fonction des cinq grades étaient 1, 4, 9, 13 et 19, réalisant une « échelle à plancher collant ». Cette échelle à « pas » irrégulier permettait de diversifier les notes totalisées par station et de

valoriser les meilleures réalisations. Ce type d'échelle n'a pas fait l'objet d'une évaluation scientifique spécifique. Le dernier item (la gestion du temps) était évalué de 1 à 5 points ; il était commun à l'ensemble des quatre stations. Chaque station était évaluée par un binôme d'examineurs ayant bénéficié d'une formation de deux jours sur les difficultés des évaluations orales et les biais de recrutement. Les huit examinateurs étaient extérieurs à l'Université d'Angers. Le même binôme évaluait tous les étudiants pour une station donnée. Les épreuves se déroulaient sur 20 demi-journées au cours desquelles 16 étudiants étaient convoqués (quatre étudiants par heure). Chaque étudiant recevait un planning qui décrivait ses horaires de passage dans les différentes stations.

Devant la complexité d'évaluer globalement une compétence, il a été décidé de « scinder » artificiellement une compétence en plusieurs items mesurables qui, à eux seuls, ne sont pas exhaustifs d'une compétence.

La station 1 évaluait les aptitudes de communication des étudiants, à partir de travaux réalisés précédemment lors des unités d'enseignement optionnelles (UEO). Le choix du sujet était libre ; les étudiants étaient évalués selon trois axes : transmission des informations au moyen d'un support visuel adapté ; organisation des idées en effectuant une présentation didactique concise ; réponses aux questions portant sur le contenu de l'exposé. Aucune compétence médicale n'était évaluée.

La station 2 visait à évaluer le travail réalisé dans l'élaboration des deux projets personnels et professionnels de l'étudiant (3PE). Les étudiants étaient invités à présenter oralement un des deux projets, après tirage au sort. Le projet n'était pas évalué en soi, mais plutôt l'utilisation des supports de réflexion, l'appropriation de la démarche d'élaboration d'un projet professionnel et la capacité à répondre aux questions.

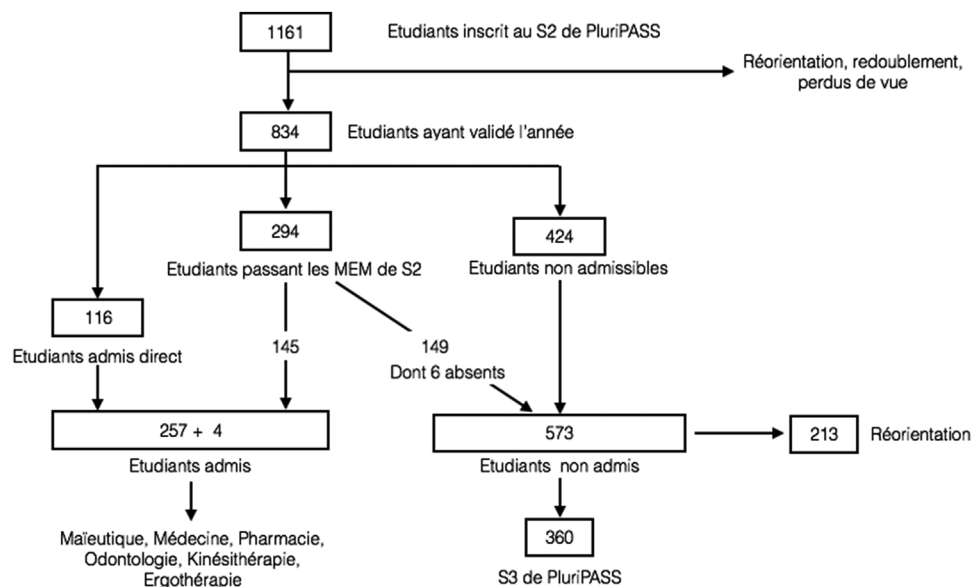


Figure. 2. Processus de l'étude.

La station 3 évaluait la réflexion autour d'une situation éthique complexe présentée sous la forme d'une vignette d'une dizaine de lignes. Elle était précédée de 10 minutes de préparation. L'étudiant devait en faire la synthèse, expliciter les points de vue sur cette situation, formuler des questions et proposer des moyens de résolution.

La station 4 portait sur l'analyse et le raisonnement scientifique à partir d'un document scientifique tiré de la grande presse (hors santé). Elle était précédée par 20 minutes de préparation. L'étudiant devait synthétiser les informations, formuler des hypothèses explicatives, formuler des questions évoquées par les documents et imaginer des perspectives. Chacune des deux dernières stations se déroulait autour de 20 scénarios différents (un par demi-journée d'épreuve) dont la difficulté avait été jugée équivalente par les responsables des stations. Ils avaient été créés par des membres de l'UA.

Deux compétences étaient transversales à chacune des stations : la réactivité et la gestion du temps.

Données

Les données démographiques collectées étaient : la date de naissance, le sexe, la nationalité, le département du baccalauréat, le type de baccalauréat et la mention, la catégorie socioprofessionnelle (CSP) des parents et l'existence ou non d'une aide boursière. Par défaut, l'âge a été défini comme celui au premier septembre de l'année scolaire. Le statut de primant ou doublant se définissait par l'absence ou non d'inscription en PACES dans les années précédentes.

Les données ont été collectées dans le logiciel APOGEE (logiciel de gestion des dossiers étudiants de l'université) à partir des inscriptions administratives des étudiants et des notes recensées par l'université.

Analyses

Les analyses ont été réalisées sur logiciel SPSS (*Statistical Package for the Social Sciences*) pour les analyses descriptives et univariées et sur le logiciel « R » pour les analyses multivariées. Les analyses ont été réalisées par le test de Mann Whitney pour échantillons indépendants dans le cas de deux groupes. Pour les analyses à plus de deux groupes, nous avons réalisé des tests ANOVA dans le cas de petits échantillons des tests non paramétriques de Kruskal-Wallis. Une correction de Benjamini Hocheberg a été appliquée pour les analyses univariées (44 tests au total), à partir du site¹.

Les analyses multivariées ont été réalisées par régression linéaire pas à pas descendante. Les variables « forcées » ont été le sexe et le statut boursier ou non. Nous avons également construit des coefficients de Cronbach à partir des scores de chaque item de chaque station. Les scores pour chaque item ont été collectés manuellement, et utilisés uniquement pour le calcul des coefficients de Cronbach. Les coefficients ont été construits par station et par compétence, ce qui nous semblait mieux respecter le principe d'unidimensionnalité de l'échelle [5] de mesure. Une revue de la littérature semble considérer un coefficient supérieur à 0,6 comme acceptable et supérieur à 0,8 comme bon, bien que ces qualificatifs n'aient pas fait l'objet de consensus [5].

L'impact des MEM a été évalué par la méthode du rang limite, qui consiste à définir le rang limite d'admission d'une filière comme étant celui du dernier étudiant à pouvoir la choisir. Ce rang est alors considéré comme celui pour lequel le *numerus clausus* aurait été atteint dans le cas où il n'y aurait pas eu de MEM, dépendant des choix mutuels des étudiants. En effet, les étudiants classés en

¹ <https://www.sdmproject.com/utilities/?show=FDR>.

Tableau I. Résultats aux mini entretiens multiples (MEM).

| | Station 1 | Station 2 | Station 3 | Station 4 | Total |
|------------|-----------|-----------|-----------|-----------|--------|
| Moyenne | 63,92 | 79,42 | 52,36 | 64,87 | 260,58 |
| Médiane | 65 | 99 | 50 | 64 | 271 |
| Écart-type | 24,26 | 27,81 | 21,01 | 17,06 | 62,47 |
| Minimum | 6 | 5 | 7 | 12 | 60 |
| Maximum | 100 | 100 | 100 | 100 | 390 |

rang utile pouvaient choisir une autre filière, tenter une nouvelle chance au concours ou refuser simplement la place. Le rang limite d'une filière est donc supérieur au *numerus clausus*. Deux difficultés étaient rencontrées dans l'analyse de l'impact des MEM: les choix des étudiants et la multiplicité des classements. Avant les MEM, un étudiant pouvait en effet être « admis direct » dans plusieurs filières, ne pas les choisir et préférer se présenter aux MEM pour tenter d'obtenir une place dans une autre filière. Après les MEM, un étudiant que les MEM classaient comme admissible dans une filière pouvait préférer passer en S3 de PluriPASS plutôt que de prendre une place dans cette filière, ou choisir une filière plutôt qu'une autre. Historiquement, les filières les plus attractives sont la médecine et l'odontologie. Si les étudiants ont une alternative ou une autre chance, la pharmacie, la maïeutique, la kinésithérapie et l'ergothérapie ne sont pas forcément choisies d'emblée par les étudiants. Ceci nous a conduit à analyser l'impact des MEM uniquement sur l'admission en médecine, le *numerus clausus* d'odontologie étant faible.

Résultats

Les résultats des MEM sont résumés dans le [tableau I](#). Si la répartition des scores aux stations 1, 3, 4 et le score total suivaient une loi normale, ce n'était pas le cas à la station 2: plus de la moitié des étudiants (152 étudiants) y ont obtenu un score entre 95 et 100.

Fidélité des MEM

Les coefficients de corrélation entre les stations allaient de 0,174 à 0,384. Ils étaient tous significatifs ($p < 0,01$). Le coefficient alpha de Cronbach global de l'épreuve était de 0,620. Les coefficients alpha de Cronbach de chacune des stations, étaient respectivement de 0,904, 0,907, 0,777 et 0,794. Nous avons analysé les coefficients selon six compétences développées par les étudiants, ce qui respectait davantage l'unidimensionnalité nécessaire à cet outil. Le coefficient rapporté pour la compétence de communication était de 0,935, celui pour la recherche documentaire de 0,928, pour la réflexion éthique 0,77, et pour le raisonnement scientifique de 0,841. Les compétences de réactivité et de gestion du temps, dont les coefficients étaient construits à partir d'items de chacune des stations, étaient respectivement de 0,612 et 0,689. Le score global des MEM (sur 400) était corrélée de façon significative avec les résultats en première année avec un coefficient de Pearson atteignant 0,400 ($p < 0,01$).

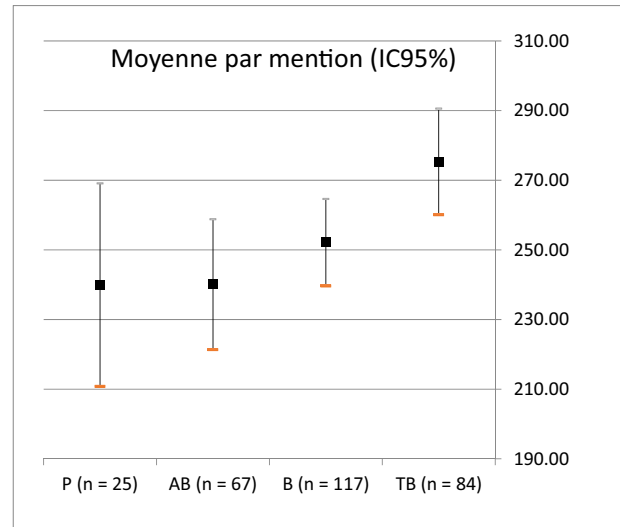


Figure 3. Moyennes et intervalles de confiance en fonction des mentions au baccalauréat pour la note totale de l'épreuve des Mini entretiens multiples (MEM).

Les étudiants ayant obtenu une mention très bien ou plus au baccalauréat ont obtenu de meilleurs scores que les autres pour la station 1 (+ 8,49 points IC95 % = [2,33 ; 14,64]), la station 3 (+ 12,45 points IC95 % = [7,23 ; 17,67]), la station 4 (+ 9,91 points IC95 % = [5,67 ; 14,16]) et pour le score total (+ 30,41 points ; IC95 % = [14,75 ; 46,12]). Il n'y avait pas de différence pour la station 2. La [figure 3](#) représente les moyennes et leurs intervalles de confiance à 95 % en fonction des mentions pour le score total à l'épreuve des MEM.

Équité des MEM

Analyse univariée

Il existait une différence en fonction du sexe des candidats ([Tableau II](#)). La différence était significative pour les stations 2 et 4 (valorisation du 3PE et analyse et raisonnement scientifique) en faveur des hommes ($n = 81$), ($p = 0,0147$ et $p = 0,020$, respectivement) par rapport aux femmes ($n = 207$). La différence de note globale des MEM était non significative après correction de Benjamini-Hochberg ($p = 0,0616$). Analysée isolément, l'analyse montrait une différence avec un écart de 27,4 points en moyenne IC95 % = [9, 2–45, 7].

Il n'a pas été retrouvé de différence significative entre les scores par les étudiants selon que la session soit du matin ou de l'après-midi.

Tableau II. Moyenne des scores obtenus aux mini entretiens multiples (MEM) en fonction du sexe.

| | Sexe | Moyenne | Écart type | <i>P</i> -value | <i>P</i> -value corrigée |
|-----------|----------|---------|------------|-----------------|--------------------------|
| Station 1 | Masculin | 64,8 | 26,30 | 0,443 | 0,750 |
| | Féminin | 63,6 | 23,47 | | |
| Station 2 | Masculin | 87,1 | 22,26 | 0,001* | 0,015 |
| | Féminin | 76,4 | 29,21 | | |
| Station 3 | Masculin | 53,4 | 22,13 | 0,727 | 0,869 |
| | Féminin | 52,0 | 20,60 | | |
| Station 4 | Masculin | 69,9 | 16,53 | 0,002* | 0,022 |
| | Féminin | 62,9 | 16,91 | | |
| Note MEM | Masculin | 275,1 | 63,54 | 0,007* | 0,062 |
| | Féminin | 247,7 | 73,71 | | |

P-value corrigée selon la méthode de Benjamini-Hochberg.

Plusieurs sujets étaient utilisés lors des MEM. Lors de la station 1, les étudiants utilisaient des sujets issus des UEO. Il n'y avait pas de différence significative selon l'UEO suivie par l'étudiant. Lors de la station 2, les étudiants utilisaient les travaux réalisés lors du 3PE. Il n'y avait pas de différence significative entre les étudiants interrogés sur le 3PE1 ou le 3PE2. Enfin, pour les stations 3 et 4, il n'existait pas de différence significative selon les scénarios utilisés.

L'âge, la CSP des parents, le fait d'être ou non boursier, le statut de primant ou redoublant, n'ont pas eu d'influence sur les résultats aux MEM.

Analyse multivariée

Le [tableau III](#) présente les coefficients de régression linéaire en fonction des variables. Le sexe et la mention « Très Bien » (respectivement coefficient 19,41 ; *p*-value = 0,011 et coefficient 31,73 ; *p*-value = 0,0325) étaient des variables associées au score total aux MEM. Le score aux épreuves écrites était également pour l'ensemble des stations (hors station 2), de même que le sexe pour la station 4. Néanmoins, les modèles obtenus à partir des variables disponibles ne prédisaient finalement qu'une faible partie de la variabilité des scores (R^2 ajusté de 0,07 à 0,1746).

Impact des MEM et de PluriPASS sur la sélection en médecine

Pour 68 places offertes en médecine après les MEM, 12 étudiants sur 13 déclassés étaient dans les 25 dernières places avant les MEM (hors un étudiant, 21^{ème} avant les MEM, qui a été classé 75^{ème} après les MEM). C'est donc les étudiants les moins bien classés avant les MEM qui ont été majoritairement exclus après les MEM. La réciproque n'était pas vraie, puisque seuls huit étudiants sur 13 finalement reçus grâce aux MEM étaient classés dans les 25 premières places au-delà du *numerus clausus* à l'issue des épreuves écrites.

Devenir des étudiants

Le [tableau IV](#) présente le devenir des étudiants exclus d'au moins une filière. Le rang limite correspondait au rang du dernier admis dans une filière après les MEM. Le

nombre d'exclus correspondait au nombre d'étudiants ayant un meilleur rang dans le classement avant les MEM et qui ont obtenu un classement au-delà de ce rang limite après les MEM. L'ergothérapie occupait une place particulière puisque toutes les places proposées en ergothérapie n'ont pas été prises. Au total, 53 étudiants ont été exclus d'au moins une filière après les MEM. Parmi eux, 20 étudiants ont été exclus d'au moins une filière sans avoir d'autre offre après les MEM et 10 étudiants ont préféré passer en S3 plutôt que de prendre une place dans une filière (+ 31 si on inclut l'ergothérapie).

Discussion

Les MEM étaient cohérents

Les coefficients de Cronbach par compétence étaient acceptables, voire très élevés. La faible corrélation entre les stations pourrait indiquer qu'elles n'évaluaient pas les mêmes compétences, ce qui était souhaité. Le coefficient global et les coefficients par stations semblaient moins pertinents puisque la condition de l'unidimensionnalité de l'évaluation n'est pas respectée. Par ailleurs, il existait une corrélation positive mais faible entre les stations, ce qui tend à penser qu'elles n'étaient pas redondantes. De même, il existait une faible corrélation entre le score total aux MEM et le score aux épreuves écrites, ce qui indiquait que les épreuves n'évaluaient pas les mêmes compétences.

Influence des mentions au bac sur les résultats aux MEM

L'avantage des étudiants ayant obtenu une mention « Très Bien » ou plus par rapport aux autres semblait confirmer l'intérêt des MEM dans le processus de sélection. Il faut rappeler que l'obtention de la mention « Très Bien » au baccalauréat peut refléter des réalités différentes et n'est pas indépendante du milieu social de l'élève. Néanmoins, nous avons montré en analyse multivariée que les étudiants avec mention « Très Bien » restaient avantagés pour le score total des MEM indépendamment de la CSP.

Tableau III. Régression linéaire pour les scores obtenus aux mini entretiens multiples (MEM). Les exclusions ont été réalisées devant des coefficients éloignés des seuils de significativité après analyse d'un premier modèle incluant les variables exclues. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

| | Note totale au MEM | | Note à la station 1 | | Note à la station 2 | Note à la station 3 | | Note à la station 4 | |
|-----------------|-----------------------|-----------------|-----------------------|-----------------|---|-----------------------|----------------|-----------------------|----------------|
| | Coefficient | <i>p</i> -value | Coefficient | <i>p</i> -value | | Coefficient | <i>p</i> value | Coefficient | <i>p</i> value |
| | R ² ajusté | 0,1746 | R ² ajusté | 0,07 | | R ² ajusté | 0,13 | R ² ajusté | 0,13 |
| Age | 0,64 | 0,323 | 0,73 | 0,497 | | -1,35 | 0,18 | -1,04 | 0,21 |
| Sexe | 19,41 | 0,011 * | 0,91 | 0,77 | | 1,4 | 0,59 | 7,51 | <0,001 *** |
| Masculin | | | | | | | | | |
| Doubleant | 2,14 | 0,77 | 2,07 | 0,48 | | 0,48 | 0,85 | 1,57 | 0,45 |
| Mention ref | ref | ref | | | | ref | ref | ref | ref |
| Passable | | | | | | | | | |
| Mention AB | 9 | 0,532 | | | | -1,81 | 0,72 | -3,11 | 0,44 |
| Mention B | 15,73 | 0,2585 | | | | -0,3 | 0,95 | 0,41 | 0,91 |
| Mention TB | 31,73 | 0,0325 | | | | 8,82 | 0,09 | 6,52 | 0,13 |
| Non boursier | -0,44 | 0,95 | -5,47 | 0,1 | Les conditions d'utilisation d'une régression linéaire ne sont pas respectées | -0,03 | 0,99 | 2,23 | 0,32 |
| CSP 1 | | | -0,28 | 0,96 | | -2,63 | 0,65 | 6,39 | 0,17 |
| CSP 2 | | | -3,01 | -0,5 | | -0,93 | 0,85 | 2,08 | 0,61 |
| CSP 3 | | | ref | Ref | | ref | ref | ref | ref |
| CSP 4 | | | -10,366 | <0,01 ** | | 1,37 | 0,68 | -0,43 | 0,87 |
| CSP 5 | | | -13,16 | 0,03 * | | 5,02 | 0,33 | -1,61 | 0,7 |
| CSP 6 | | | -3,53 | 0,48 | | -8,45 | 0,04 * | -2,14 | 0,53 |
| CSP 8 | | | 2,38 | 0,88 | | 11,46 | 0,42 | -5,01 | 0,66 |
| CSP 9 | | | -11,28 | 0,52 | | -1,3 | 0,93 | 2,57 | 0,83 |
| Note en licence | 19,87 | <0,001 *** | 5,84 | <0,001 *** | | 5,1 | <0,001 *** | 3,2 | <0,001 *** |

Mention AB/B/TB : Assez bien, Bien, Très bien ou plus au baccalauréat.

CSP : Catégorie socio professionnelle.

Tableau IV. Devenir des étudiants refusés dans une filière.

| Étudiants refusés en | Effectif (exclus) | % du NC | Refus de place offerte dans une autre filière pour passer en S3 | Pas d'autre offre | Prise de place en | | | | | | Rang limite |
|----------------------|-------------------|---------|---|-------------------|-------------------|---|----|---|---|---|-------------|
| | | | | | M | K | MA | O | P | E | |
| Médecine (M) | 13 | 10 % | 4 | 3 | | 5 | 0 | 0 | 1 | 0 | 69 |
| Kinésithérapie (K) | 16 | 57 % | 4 | 8 | 0 | | 1 | 0 | 2 | 1 | 104 |
| Maïeutique (MA) | 16 | 84 % | 5 | 4 | 1 | 3 | | 0 | 2 | 1 | 55 |
| Odontologie (O) | 17 | 113 % | 3 | 1 | 8 | 3 | 0 | | 2 | 0 | 43 |
| Pharmacie (P) | 14 | 25 % | 2 | 12 | 0 | 0 | 0 | 0 | | 0 | 124 |
| Ergothérapie (E) | 0 | | 31 | - | - | - | - | - | - | - | 80 |

Lecture : 13 étudiants ont été refusés en médecine. Quatre ont fait le choix de passer en S3 malgré des places proposées dans d'autres filières, 3 n'avaient aucune autre place proposée, 5 sont allés en kinésithérapie et 1 en pharmacie.

NC : Numérus clausus.

Les MEM sont-ils sexistes ?

En analyse univariée, les hommes ont obtenu de meilleurs résultats que les femmes pour les stations 2 et 4 et pour le score total. Cet avantage existait également en analyse multivariée pour le score total. Ce résultat n'est pas rapporté dans la littérature [3,4,6,7]. Lors de leur formation, les jurés avaient été sensibilisés à un risque de biais de notation lié au sexe. Il est possible qu'on attribue plus facilement une compétence à un sexe ; mais il se peut aussi que les épreuves, telles qu'elles ont été construites et mises en œuvre, fassent ressortir des différences liées à l'éducation. Ce constat doit surtout inciter à travailler cette question en profondeur, tant lors de la conception des MEM (épreuves, grilles de critères) que dans leur réalisation (formation des jurés, déroulement).

Les MEM étaient équitables

En analyse univariée ou multivariée, aucun avantage ni désavantage n'a été montré pour les redoublants dans l'épreuve des MEM. Les analyses univariées et multivariées ont semblé montrer que les MEM étaient indépendants du moment de la session dans la journée, de l'âge, de l'aide boursière, de la CSP des parents.

Le rôle des différents scénarios

L'utilisation de différents scénarios (station 3 et 4), ou de support d'UEO (station 1) ou de support de 3PE (station 2) n'a pas entraîné de différence statistiquement significative. Des analyses de généralisabilité ou des modèles de Rasch pourraient être pertinents pour analyser l'impact des scénarios sur les scores. En effet, la plupart des travaux publiés ne portaient pas sur différents scénarios pour une même station, hors les études de Zaidi *et al.* [8] (dont la construction a été réalisée *a posteriori* de l'épreuve et sur un très faible effectif) et celle de Roberts *et al.* [9]. Cette étude retrouve que la variance du score pour une station est influencée à 22 % par l'étudiant, à 14 % par la sévérité de l'examinateur, à 2 % par la difficulté du scénario, à 8 % par l'intérêt de l'examinateur pour le scénario. Les 54 % restants ne sont pas attribuables et l'intérêt de l'étudiant pour le sujet du scénario ne peut pas être pris en compte de manière quantifiable. À l'UA, la sévérité de l'examinateur n'intervient pas, puisque tous les étudiants passent devant les mêmes binômes d'examineurs et que le concours produit un classement relatif et non absolu. Ainsi, si la sévérité d'un examinateur intervient bien sur le score absolu d'un étudiant, elle intervient également pour tous les autres. Elle n'intervient donc pas sur le classement relatif d'un étudiant vis à vis des autres.

Impact des MEM

Les MEM ont eu un impact certain sur les classements d'admission dans les filières sélectives, classements que nous avons choisi d'analyser par la méthode des rangs limites. Pour chaque filière, 10 à 20 places dans le *numerus*

clausus ont été réattribuées en fonction des MEM. Cinquante trois étudiants différents sur 294 ont été exclus d'une filière après les MEM et autant ont obtenu une place dans une filière grâce aux MEM. Au total, les MEM ont donc modifié l'avenir d'environ un tiers (106/294) de la population qui y a concouru.

Limites de l'étude

La première limite de la présente étude est son caractère exploratoire. Le choix a été fait de chercher d'abord à faire ressortir des différences avant de les explorer et de les traiter. Il a été choisi de corriger la multiplicité des tests en maîtrisant le *false discovery rate* (FDR) par la méthode de Benjamini-Hochberg. Il est important de prendre ces résultats comme des indications et non comme des vérités absolues, et de les valider sur de plus grands échantillons.

Cette étude a été réalisée d'abord comme un outil d'aide à la décision pour le comité de pilotage de PluriPASS et pour les comités de rédaction des épreuves. Un de ses buts était d'attirer l'attention sur des éventuels problèmes, par exemple sur le sexisme, plus que d'aller au fond d'un unique problème. Néanmoins, il faut également souligner qu'il manque certaines comparaisons, par exemple en terme socio démographique. En France, il est prohibé (hors cas exceptionnel) de recueillir l'origine ethnique des étudiants, qui pourrait être un biais dans la notation des étudiants.

Cette étude explore, par exemple, la fidélité des épreuves par le coefficient alpha de Crombach comme principal outil, sans utiliser de méthodes plus fines telles qu'une étude de généralisabilité et une étude par méthode de Rasch dans le cadre du fonctionnement différentiel des items. Ces deux méthodes pourraient chacune faire l'objet de travaux spécifiques qu'il n'a pas été possible de réaliser lors de la première année de PluriPASS.

La validité de l'épreuve n'est également pas abordée spécifiquement par cette étude. La validité fait référence à la capacité d'une épreuve à démontrer un élément externe à l'épreuve, ici une compétence de l'étudiant. L'absence de matériel de comparaison est la principale raison de l'absence d'étude spécifique de la validité de l'épreuve : d'une part, parce qu'il n'existe en France que très peu d'épreuves non purement académiques ; d'autre part, parce que l'étude a eu lieu pendant les épreuves, ce qui ne permet pas de comparaison avec des données ultérieures, par exemple au moment où les étudiants se retrouvent confrontés aux patients, à des réflexions éthiques et scientifiques.

Comparaison à d'autres travaux

Particularités de l'expérience d'Angers

L'introduction des MEM était une première en France. Les MEM ont déjà été évalués dans le contexte francophone [10] mais peu d'études [8,9,11] rapportent l'utilisation des MEM dans un processus de sélection.

Néanmoins, les MEM de l'UA présentaient un design différent de celui des MEM utilisés ailleurs. Ils comportaient quatre épreuves contre 10 en général. Chaque grille évaluait 5 + 1 compétences à Angers contre trois à quatre ailleurs. L'échelle de notation « à plancher collant » était également originale, par rapport aux échelles de type Likert couramment utilisées. Les étudiants passant les MEM à Angers venaient d'obtenir le baccalauréat (un an auparavant) tandis que la plupart des articles décrivant des MEM s'intéressaient à des étudiants de grade licence ou master. Les MEM de l'UA avaient également la particularité, pour les deux premières stations, de renvoyer à des travaux réalisés en amont par les étudiants. Cette particularité s'expliquait par la volonté de valoriser des enseignements innovants au sein d'une première année d'études de santé (3PE et UEO). Ces différences nécessitent de la prudence lors de la comparaison avec d'autres travaux.

La corrélation entre les scores obtenus pour chaque station montrait qu'un étudiant réussissant bien une station réussissait bien les autres, bien que de nombreux facteurs influencent cette corrélation. Les études qui ont analysé cette corrélation entre les MEM retrouvent des corrélations pour la plupart significatives mais faibles. L'expérience de Dundee retrouve des corrélations allant de 0,057 à 0,363 et de -0,061 à 0,308 [11] (années 2009 et 2010). La meilleure corrélation entre deux stations, obtenue à l'Université Mac Master, atteint 0,370 [1]. Ces corrélations sont significatives mais faibles, ce qui peut indiquer, dans la littérature comme dans notre étude, que les stations ne sont pas redondantes.

La difficulté essentielle dans l'utilisation du coefficient de Cronbach est de délimiter le périmètre de la compétence analysée (ou de définir quels items doivent être intégrés au calcul, pour que le coefficient n'analyse qu'une compétence). « La cohérence interne est une condition nécessaire, mais non-suffisante de l'homogénéité des items d'une échelle, cette dernière impliquant l'unidimensionnalité de l'échelle en question » [5]. Par exemple, l'expérience de Dundee en 2009 et 2010 [11] a été construite à partir de 10 stations, évaluant (par trois sous-scores) une, deux ou trois des six compétences prédéfinies en amont (communication, raisonnement logique et esprit critique, raisonnement éthique, motivation, intégrité, travail en équipe). Un premier coefficient de Cronbach a été calculé sur les scores de chaque station, et s'élevait à 0,65. Une analyse des scores par compétences a été ensuite réalisée : elle retrouvait des coefficients acceptables pour la compétence « communication » en 2009 et 2010 ($\alpha = 0,69$ et $0,70$) et faibles pour toutes les autres (α allant de $0,10$ à $0,45$).

Dans notre étude, le coefficient global de l'épreuve s'élevait à 0,620. Mais ce coefficient ne semblait pas pertinent puisque construit à partir d'items de différentes compétences. Nous avons par contre retrouvé des coefficients de compétences élevés quand elles n'étaient évaluées que dans une seule station (et donc par un seul binôme de jurés) et acceptables quand elles étaient évaluées dans plusieurs. Les forts coefficients attribués

dans une seule station témoignent peut-être d'un « biais de halo » subi par les jurés (attribuer le score d'un item en fonction des autres items).

Par ailleurs, aucune étude ne retrouve de différence significative selon l'âge, la CSP des parents, le moment de la journée [3,4]. Les corrélations avec les épreuves académiques sont positives mais faibles, dans notre étude comme dans la littérature, ce qui peut démontrer que les épreuves ne sont pas redondantes [9,11].

Conclusion

Ce premier travail sur les MEM au sein de l'expérimentation PluriPASS démontre qu'ils ont un intérêt pédagogique et sélectif, et qu'ils ont une cohérence interne et externe. Les MEM doivent se placer dans une stratégie pédagogique et dans un processus de sélection, ancrés dans leur environnement. Néanmoins, ces épreuves nécessitent une attention particulière dans leur conception, leur réalisation et dans leur analyse pour en garantir l'équité. Dans ce sens, d'autres études sont nécessaires pour évaluer d'autres variables non étudiées, par exemple les différences socio-culturelles entre les étudiants. Il semble également pertinent de poursuivre les analyses sur l'impact des différents scénarios, par exemple par des études de généralisabilité ou utilisant un modèle de RASCH.

Contributions

Antoine Jacquet a participé à la conception du protocole de recherche, au recueil des données, à l'interprétation des résultats, à l'analyse statistique et à l'écriture du manuscrit. Jérémie Riou a participé à la conception du protocole de recherche, à l'analyse statistique et à l'écriture du manuscrit. Elisabeth Letertre a participé au recueil et à l'analyse des données. Catherine Passirani a participé au recueil des données, à l'interprétation et à l'analyse des résultats, ainsi qu'à l'écriture du manuscrit. Jean-Paul Saint-André a participé à la conception du protocole de recherche, à l'interprétation des résultats et à l'écriture du manuscrit.

Approbation éthique

Ces travaux n'ont pas fait l'objet d'avis d'une commission d'éthique spécifique. Néanmoins, l'utilisation des données démographiques et des résultats des étudiants de façon anonyme a été plusieurs fois présentée en comité d'évaluation extérieur à PluriPASS (Comité composé de membres universitaires extérieurs à l'Université d'Angers, et de membres de la société civile – parlementaire, conseil économique social et environnemental, associations).

Conflits d'intérêts

Les auteurs ne déclarent pas de liens d'intérêt en lien avec le contenu de cet article.

Références

1. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: The multiple mini interview. *Med Educ* 2004;38:314-326.
2. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The objective structured clinical examination (OSCE): AMEE Guide No.81. Part I: An historical and theoretical perspective. *Med Teach* 2013;35:437-446.
3. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The objective structured clinical examination (OSCE): AMEE Guide No.81. Part II: An historical and theoretical perspective. *Med Teach* 2013;35:1447-1463.
4. Pau A, Jeevaratnam K, Chen YS, Fall AA, Khoo C, Nadarajah VD. The multiple mini-interview (MMI) for student selection in health professions training-A systematic review. *Med Teach* 2013;35:1027-1041.
5. Laveault D. Soixante ans de bons et mauvais usages du alpha de Cronbach. *Mes Eval Educ* 2012;35:1-7.
6. Kelly ME, Dowell J, Husbands A, Newell J, O'Flynn S, Kropmans T, *et al.* The fairness, predictive validity and acceptability of multiple mini-interview in an internationally diverse student population, a mixed methods study. *BMC Med Educ* 2014;14:261.
7. Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Med Educ* 2009;43:573-579.
8. Zaidi NB, Swoboda C, Wang LL, Manuel RD. Variance in attributes assessed by the multiple mini-interview. *Med Teach* 2014;36:794-798.
9. Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, *et al.* Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Med Educ* 2008;42:396-404.
10. Renaud JS, Cantat A, Lakhil S, Bourget M, Saint-Onge C. Sélection des candidats en médecine : validité prédictive des mini entretiens multiples en contexte francophone. *Pédagogie Médicale* 2016;17:7-21.
11. Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. The multiple mini-interview in the UK Context: 3 years of experience at Dundee. *Med Teach* 2012;34:297-304.

Citation de l'article : Jacquet A., Riou J., Letertre E., Passirani C., Saint-André J.-P., Fidélité, équité et impact des mini entretiens multiples dans le processus de sélection des étudiants en santé en France. *Pédagogie Médicale* 2017;18:179-188