

Epreuves multidisciplinaires dans les études médicales : Apport de l'analyse docimologique

Multidisciplinary exams in medical studies : Interest of docimologic analysis

Naouel Ben Salah, Issam Salouage, Zied Zein El Abidine Hatira, Rim Goucha, Ahmed Meherzi, Kalthoum Kallel

Faculté de Médecine de Tunis/ Université Tunis El Manar

RÉSUMÉ

Introduction : La docimologie a permis le développement de procédés évaluatifs garantissant des épreuves valides, fiables et objectives. Elle a été adoptée à la faculté de médecine de Tunis depuis 2007. Le but de ce travail était d'analyser des résultats de l'étude docimologique des épreuves d'hématologie-oncologie, d'évaluer l'intérêt de cette analyse dans l'élaboration des épreuves et la construction d'une banque d'items et de proposer des mesures correctives en vue d'améliorer les épreuves.

Méthodes : Nous avons analysé les notes, préalablement saisies sur Excel, des épreuves d'hématologie-oncologie, du DCEM1, de 2008-2009 à 2013-2014. L'épreuve comporte 4 disciplines (hématologie, oncologie, génétique humaine et anatomie pathologique). Nous avons calculé les paramètres docimologiques permettant une analyse globale, par discipline et par item.

Résultat : 3281 copies et 1004 questions ont été analysées. Le taux moyen de réussite par an était de $91,54\% \pm 7,12$. Le meilleur taux moyen de réussite était noté pour l'hématologie ($80,51\% \pm 10,18$). Le taux le plus faible ($51,61\% \pm 23,76$) était noté pour l'anatomie pathologique. Le taux moyen d'étudiants ayant réussi l'épreuve sans réussir l'hématologie était de 5,36%. Il était de 42,29% pour l'anatomie pathologique. L'indice moyen de difficulté par épreuve était de $0,57 \pm 0,05$. Nous avons noté 38,04% questions faciles et 19,02% questions difficiles. L'indice moyen de discrimination était de $0,25 \pm 0,02$. La discrimination était très bonne pour 20,51% des items, et bonne pour 17,13%. Le coefficient Alpha de Cronbach moyen était de $0,84 \pm 0,03$ témoignant d'une bonne homogénéité interne.

Conclusion : Notre étude a permis d'évaluer objectivement la place des disciplines dites contributives dans l'évaluation multidisciplinaire et de montrer l'intérêt de l'intégration des questions permettant une meilleure évaluation. L'analyse des questions de cette épreuve en concertation avec les enseignants permettrait une interprétation plus approfondie de ces résultats.

Mots-clés

Docimologie, Evaluation, Pédagogie, Sciences Médicales

SUMMARY

Background : Docimology has allowed the development of evaluative processes assuring valid, reliable and objective assessments. It was adopted within the faculty of Medicine of Tunis since 2007.

The aim of this study was to analyze the docimological survey results of hematology-oncology exams, to evaluate the interest of this analysis in the elaboration of exams and the construction of an item bank, and propose some corrections in order to improve assessment.

Methods : We have analyzed the hematology-oncology exams of SCMS1 (Second cycle of Medical Studies 1) from educational year 2008-2009 to 2013-2014. The data input was already done with Excel. The test includes 4 disciplines (Hematology, Oncology, Genetics and the Anatomic Pathology). We have calculated docimological parameters allowing global analysis, by discipline and by item.

Results : A total of 3281 papers and 1004 questions were analyzed. The average success rate per year was $91,54\% \pm 7,12$. The highest average success rate was found in hematology ($80,51\% \pm 10,18$). The lowest rate was found in the anatomic pathology ($51,61\% \pm 23,76$). The average rate of students succeeding the test without having average note in hematology was 5,36%. It was 42,29% in the anatomic pathology. Average difficulty index was $0,57 \pm 0,05$. Items analysis showed that 38,04% were easy and 19,02% were difficult. Average discrimination index was $0,25 \pm 0,02$. Discrimination was very good in 20,51% of items and good in 17,13%. Useless and bad discrimination items were about 40,53%. The average of Cronbach Alpha coefficient was $0,84 \pm 0,03$, showing a good internal-consistency.

Conclusion : This study allowed an objective evaluation of "contributive disciplines" in multidisciplinary evaluation and showed the interest of integrating questions. Question analysis with teachers would be important to reevaluate and improve these items.

Key-words

Docimology, Assessment, pedagogy, Medical Sciences

INTRODUCTION

La docimologie a été fondée au début des années 1920 par deux scientifiques français, Henri Piéron et Henri Laugier (1). Ils ont réalisé, en 1922, les premières études selon les méthodes objectives de la science expérimentale (2). La docimologie ou « science des examens » a pour objectif d'étudier l'examen dans toutes ses dimensions et d'en juger la qualité (3,4). Elle a permis de cerner les biais liés à l'évaluation, et de publier des recommandations et des « guidelines » suivis au sein des plus grandes universités (5-7).

La docimologie a été introduite à la Faculté de Médecine de Tunis depuis 2007, par la mise en place d'un processus de vérification des notes et d'évaluation des épreuves par l'analyse docimologique.

L'objectif de ce travail était de rapporter la démarche docimologique adoptée à la FMT en prenant comme exemple l'épreuve d'hématologie oncologie enseignée en première année du deuxième cycle des études médicales (DCEM1) à la Faculté de Médecine de Tunis et d'en dégager les points forts et les points faibles.

MÉTHODES

Il s'agit d'une étude descriptive rétrospective. Elle a consisté en une analyse docimologique deconcerné toutes les épreuves d'hématologie-oncologie des trois sessions sur une période de 6 années universitaires, de 2008-2009 à 2013-2014.

Présentation de l'épreuve

Le module Hématologie-Oncologie est enseigné en DCEM1. Il évalue 4 disciplines dont la notation était connue par l'étudiant. L'hématologie est notée sur 15 points, elle compte 45 questions. L'oncologie est notée sur 5 points. La génétique humaine (3 questions) est notée sur 3 points. L'anatomie pathologique (4 questions) est notée sur 2 points.

Saisie des notes et vérification :

Après correction et calcul manuel des moyennes. Toutes les notes ont été saisies sur Excel® avant la levée d'anonymat. Les totaux et sous totaux de chaque discipline ont été calculés ainsi que le total de l'épreuve et la moyenne générale. Les erreurs de saisie et les erreurs du total manuel ont été vérifiées.

Analyse docimologique

Analyse globale : Dans cette étape ont été analysés les taux de réussite, les moyennes maximales et minimales.

Analyse par discipline : Nous avons comparé le taux de réussite à l'épreuve et le taux des étudiants ayant réussi l'épreuve sans réussir la discipline.

Analyse par item : Nous avons calculé pour chaque item :

- Pourcentage des questions notées zéro : Nous nous sommes intéressés aux questions dont le pourcentage de zéro était supérieur ou égal à 30% (8).
- Pourcentage des questions avec la note maximale : Nous nous sommes intéressés aux questions dont le pourcentage était supérieur ou égal à 80% (8).
- Indice de difficulté : C'est le pourcentage de réussite par question. Il définit la difficulté de la question. Il est représenté par le symbole p , il est calculé par la formule :

$$p = \frac{\text{Moyenne de la question}}{\text{Barème de la question}}$$

L'indice de difficulté peut prendre des valeurs allant de 0 à 1.

L'indice de difficulté est interprété de la manière suivante (8):

- [0 - 0,3[→ question difficile
- [0,3 - 0,7[→ question à difficulté acceptable
- [0,7 - 1] → question facile

Nous calculons l'indice de difficulté moyen, par session et par discipline.

- Indice de discrimination : Il correspond à la capacité de l'item à départager les « forts » des « faibles ». Il correspond à la différence entre l'indice de difficulté d'un item pour le groupe dit « fort » : $p+$ (30% premiers selon la moyenne générale) et l'indice de difficulté pour le groupe dit « faible » : $p-$ (30% derniers).

$$D = (p+) - (p-)$$

L'indice de discrimination varie de -1 à 1. Il est interprété selon Ebel (8) : Nous avons calculé l'indice de discrimination pour chaque item. L'indice de discrimination par discipline n'a été appliqué qu'à l'hématologie, les autres disciplines ayant un faible nombre de questions.

L'indice de discrimination global correspond à la moyenne des indices par items :

$$D_{global} = \frac{\text{Somme des indices } D \text{ par item}}{\text{Nombre d'items}}$$

Evaluation de l'ensemble de l'épreuve : Coefficient Alpha de Cronbach

Il est appelé encore indice de fidélité ou de fiabilité. Il permet d'évaluer l'homogénéité interne de l'épreuve. Le coefficient alpha se calcule en appliquant la formule suivante :

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum \sigma_j^2}{\sigma_T^2} \right]$$

k = nombre de questions

σ_j^2 = variance des scores à une question j

σ_T^2 = variance des scores à l'ensemble de l'épreuve ou du test

Il varie de 0 à 1. Dans la pratique, on considère généralement que l'homogénéité de l'instrument est satisfaisante lorsque la valeur du coefficient est au moins égale à 0,8 (8).

Interprétation des valeurs du coefficient de Cronbach selon Goerge et Mallery (9):

- [1 - 0,90] : Excellent
-]0,9 - 0,8] : bon
-]0,8 - 0,7] : acceptable.
- <0,7 : non acceptable

Ensuite nous avons calculé l'alpha global comme suit :

$$\alpha_{global} = \frac{\text{Somme des Coefficients Alpha de Cronbach par épreuve}}{\text{Nombre d'épreuves}}$$

Analyse statistique

Les résultats ont été exprimés en pourcentage et en moyenne $\pm$ écart type (minimum - maximum). Pour l'étude descriptive, les fréquences simples ont été calculées pour les variables qualitatives et les moyennes. Les écarts types ont été calculés pour les variables quantitatives. Pour l'étude analytique, la comparaison des moyennes a été effectuée au moyen du test T de STUDENT. La comparaison des pourcentages sur séries indépendantes a été effectuée par le test Chi-deux de Pearson. Le seuil

de signification de p a été fixé à 0,05.

RESULTATS

Analyse globale

- Données générales

Durant la période d'étude nous avons analysé : 3281 copies, 1004 questions et 193579 données

- Taux de réussite

Le taux moyen de réussite par an était de $91,54\% \pm 7,12$ ($80,97\% - 96,77\%$). Le taux moyen de réussite par session était $72,90\% \pm 15,26$ ($42,17\% - 95,78\%$).

Les moyennes maximales et minimales par session sont représentées dans le tableau 1

Tableau 1: Les moyennes maximales, minimales et la moyenne des moyennes par session

		Année universitaire						
		Session 2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	
Note Maximale (/20)	Janvier	16.20	16.93	ND	16.46	17.86	16.87	
	Mai/Juin	18.00	18.73	17.13	18.47	17.47	18.20	
	Juillet	15.33	14.73	16.47	18.60	13.53	15.53	
Note minimale (/20)	Janvier	2.50	4.40	ND	2.60	1.06	2.67	
	Mai/Juin	5.26	4.33	4.53	1.93	4.20	1.47	
	Juillet	4.67	3.33	4.87	5.93	3.60	2.00	
Moyenne des moyennes (/20)	Janvier	12,06	10,81	ND	10,06	12,52	10,56	
	Mai/Juin	12,01	13,44	11,66	11,49	12,10	10,52	
	Juillet	10,32	11,00	11,81	13,05	9,50	10,01	

ND : non déterminé

Analyse par discipline

Les taux de réussite les plus élevés ont concerné l'hématologie : $80,5\%$ pour les sessions principales et $74,4\%$ pour les sessions de rattrapage. Les taux de réussite les moins élevés ont concerné l'anatomopathologie. Les taux de réussite selon les disciplines ainsi que les taux d'étudiants qui ont réussi l'épreuve sans avoir réussi la discipline sont représentés dans le tableau 2.

Analyse par item

Questions notées zéro : Le taux des questions avec 30% ou plus de zéro, était de $45,32\%$. Ce taux de question avec au moins 30% de zéro selon les disciplines est représenté

dans le tableau 3.

Tableau 2 : Taux de réussite par discipline

Discipline	Session principale	Session de rattrapage	Taux de réussite de l'épreuve sans réussir la discipline
Hématologie			
Moyenne %	80,51 ± 10,18	74,40 ± 14,81	5,36 ± 8,20
Extrêmes %	(65,54 – 94,92)	(50,77 - 93,97)	(0 -29,55)
Oncologie			
Moyenne %	68,08 ± 26,31	65,47 ± 33,58	24,99 ± 26,28
Extrêmes %	(16,13 – 95,34)	(9,64 - 96,39)	(1,93 -80)
Anatomopathologie			
Moyenne %	51,61 ± 23,76	44,70 ± 14,98	42,29 ± 20,91
Extrêmes %	(20,22 – 88,31)	(19,9 - 60,47)	(3,21 -69,04)
Génétique			
Moyenne %	61,55 ± 17,07	60,07 ± 20,60	27,17 ± 14,02
Extrêmes %	(36,60 – 90,64)	(25,3 - 86,05)	(3,2 -60)

Tableau 3 : Taux de questions avec 30% ou plus de zéro par discipline

	Taux de zéro ≥ 30%
Hématologie	45,33
Oncologie	35,39
Génétique	44,44
Anatomie pathologique	62,68
Globalement	45,32

Questions avec la note maximale : Le taux des questions avec 80% ou plus de notes maximales, était de 18,76% ± 6,54 (9,83% - 33,33%). Ce taux selon les disciplines est représenté dans le tableau 4.

Indice de difficulté :

- L'indice moyen de difficulté par session était de 0,57 ± 0,05 (0,5 – 0,68).
- La comparaison des indices moyens de difficulté entre les différentes sessions n'a pas montré de différence significative.
- Selon la difficulté, les questions étaient faciles dans 38,04%, difficiles dans 19,04% des cas et de difficulté acceptable dans 42,92% des cas. L'étude statistique n'a pas montré de différence statistique entre les disciplines selon les indices de difficulté.

Indice de discrimination : L'indice moyen de discrimination était de 0,25 ± 0,02 (0,21 – 0,31) pour l'ensemble des épreuves, et de 0,26 ± 0,03 (0,22 – 0,33) pour la discipline d'hématologie. Une seule épreuve avait une bonne

discrimination. La répartition des questions selon leur discrimination est représentée dans le tableau 5.

Tableau 4 : Taux de questions avec au moins 80% de note maximale par discipline

	Taux de questions ayant eu la note maximale ≥80%
Hématologie	22,86
Oncologie	4,42
Génétique	9,52
Anatomie pathologique	2,98
Globalement	18,76

Tableau 5 : Répartition des questions selon l'indice de discrimination

Répartition des questions selon leur indice de discrimination		
Discrimination	Hématologie	Toutes disciplines confondues
Très bonne (%)	21,81	20,51
Bonne (%)	19,71	17,13
Moyenne (%)	23,25	21,81
Mauvaise (%)	21,68	24,90
Sans utilité (%)	13,53	15,63

Coefficient Alpha de Cronbach : Le coefficient Alpha de Cronbach moyen était de 0,84 ± 0,03 (0,77 – 0,88), correspondant à une bonne homogénéité interne.

Parmi les 17 sessions, deux avaient une homogénéité interne acceptable. Les autres sessions avaient une bonne homogénéité.

Nous n'avons pas relevé de variabilité entre les différentes sessions.

DISCUSSION

Dans cette étude nous avons effectué une analyse docimologique d'une épreuve multidisciplinaire, l'hématologie oncologie. L'objectif était d'effectuer une évaluation globale par discipline et par item pour une étude objective de l'épreuve. Le taux moyen de réussite annuel était élevé. L'indice moyen de difficulté par session était de 0,57. L'indice moyen de discrimination était de 0,25. Le coefficient Alpha de Cronbach moyen était bon. Les

résultats ont montré que les disciplines contributives et particulièrement l'anatomopathologie seraient négligées par les étudiants avec des taux élevés d'étudiants qui réussissent le module sans réussir la discipline.

Analyse docimologique globale

A notre connaissance, il n'y a pas eu d'études docimologiques portant sur des épreuves d'hématologie oncologie. L'analyse globale a été faite comparativement à des études concernant des épreuves de sciences médicales (10,11). Les données de la littérature montrent une hétérogénéité des taux de réussite (11-13).

Analyse par discipline

Le taux de réussite de l'épreuve suivait sensiblement celui de l'hématologie. Ceci est lié au poids important de l'hématologie par rapport aux autres disciplines avec un désistement volontaire s'investir dans la préparation des disciplines à « faible » notation.

En plus de la faible pondération de l'anatomie pathologique il y aurait eu un désintérêt par rapport à la discipline. Une étude faite à la faculté de médecine d'Amiens, avait montré que 89% des étudiants estimaient leurs connaissances en anatomie pathologique insuffisantes, 53 % jugeant la formation théorique médiocre. Pourtant, 85% des étudiants trouvaient utiles les travaux pratiques d'anatomie pathologique et 93% reconnaissaient que l'anatomie pathologique jouait un rôle majeur dans la médecine moderne. Seuls 38% estimaient qu'il s'agissait d'une discipline intéressante (14). Une autre étude a montré que cette spécialité ne paraissait pas beaucoup intéresser les jeunes étudiants (15).

Nos résultats ainsi que la revue de la littérature, soulèvent la problématique de l'enseignement des disciplines « contributives », leur évaluation de se heurte à un désintérêt puisque les étudiants réussissent bien l'épreuve sans réussir ces disciplines. La spécificité de ces disciplines impose une réflexion approfondie au sein des comités en charge. C'est là tout l'intérêt de l'intégration des questions. L'épreuve, n'évaluant pas chaque discipline à part, elle propose un problème dont la résolution nécessite la maîtrise des différents volets de ces disciplines. Ceci aurait l'avantage d'éviter à l'étudiant de préparer une discipline selon sa pondération dans l'épreuve. Ceci permettra une meilleure évaluation et une meilleure formation. En effet, l'évaluation de ces disciplines « contributives » est aujourd'hui biaisée par le manque de

préparation. La moyenne obtenue ne reflète pas la qualité de l'enseignement ni le niveau de connaissances.

Ces données soulèvent, aussi, le problème du niveau de connaissances de nos médecins qui parviennent à réussir leurs études sans acquérir un minimum de connaissance dans des disciplines comme l'anatomie pathologique ou la génétique humaine.

Cette réflexion devrait concerner tous les enseignements avec des « disciplines contributives ».

Analyse par item

Le taux de questions notées zéro a été globalement de 45,32%.

Une question notée zéro pour une proportion importante des étudiants pourrait être liée à plusieurs facteurs dont la difficulté réelle, l'ambiguïté ou l'omission faute de temps. Notre étude devrait être complétée par une évaluation de la formulation des questions elles-mêmes.

Le Taux des questions ayant eu la note maximale renseigne sur les questions où 80% au moins des étudiants, ont eu la note maximale. Il s'agit donc d'un indicateur de maîtrise des questions.

L'obtention de la note maximale pourrait être expliquée par plusieurs hypothèses, dont on citera essentiellement : question facile, question précédemment parue dans les annales, objectif d'enseignement atteint.

L'interprétation de ce taux seul n'est pas assez informative. Elle doit être couplée à d'autres indices telle la difficulté ou la discrimination. En effet, ce taux ne reflète pas la qualité des étudiants ayant répondu à la question : s'agit-il des forts ou des faibles ? Discrimine-t-elle ou pas ces deux groupes ?

Indice de difficulté

Notre étude a montré que les items de difficulté acceptable. Ces proportions s'accordaient avec les recommandations de l'Université de Lausanne qui stipulent qu'un bon équilibre comprend 40% d'items faciles auxquelles tous les étudiants peuvent répondre, 50% d'items qui exigent raisonnement et réflexion et 10% d'items complexes auxquelles seuls les meilleurs étudiants peuvent répondre (16).

Un indice de difficulté qui tend vers 1 pourrait être expliqué par plusieurs facteurs dont : l'objectif pédagogique est atteint, question réellement facile ou question parue dans les annales de préparation (17). L'intervention de ce facteur serait minime lors de la préparation de l'épreuve.

Cependant un « épuisement » des questions peut être observé avec le temps.

Cet indice tend au contraire vers zéro dans plusieurs situations. La difficulté peut être liée à l'enseignement lui-même, à la formulation de la question (question réellement difficile ou ambiguë) ou suite un problème lié à la correction. La non-réponse (taux du non répondant par item) n'a pas été évaluée dans notre étude, vu que les « non-réponses » ne sont pas analysées. En effet le zéro peut correspondre à une question avec une réponse erronée ou à une « non-réponse ». La saisie des « non-réponses » pourrait améliorer l'évaluation de ce facteur. En effet une non-réponse peut simplement être liée au facteur temps.

Indice de discrimination

Le caractère discriminatif est défini par la capacité de l'item, à départager les « forts » des « faibles » par rapport à la note générale de l'épreuve.

Les épreuves étaient moyennement discriminantes. Au niveau disciplinaire, seule l'hématologie a été analysée. En effet, le calcul de l'indice de discrimination doit se faire à partir d'un nombre suffisant d'items (30 ou plus). Lorsque le nombre d'items est restreint, l'indice de discrimination est artificiellement surestimé du fait que chaque item compte pour une proportion importante du total (8,18). Dans notre étude seule l'hématologie répondait à ce critère. Son indice de discrimination était de $0,26 \pm 0,03$, ce qui est sensiblement similaire à l'indice global.

La discrimination n'est pas un objectif dans ce type d'épreuve. Elle trouve son intérêt plutôt dans les concours.

Homogénéité interne - Coefficient Alpha de Cronbach

La fiabilité, ou fidélité d'une épreuve, désigne sa capacité à mener à des résultats similaires d'une évaluation à une autre. C'est l'un des trois critères majeurs dans la littérature docimologique désignant une épreuve de qualité. En effet, la fidélité des résultats est essentielle pour atteindre la validité. Sans fidélité, les résultats ne peuvent être ni pertinents, ni utiles (8).

Créé par Lee Cronbach en 1951, cet indice statistique est la mesure de fidélité la plus utilisée dans la littérature docimologique (19-21). Le coefficient alpha de Cronbach varie de 0 à 1. Plusieurs niveaux de fidélité recommandée ont été proposés selon le champ d'application. Le test est généralement considéré satisfaisant ou acceptable dès lors que le coefficient alpha de Cronbach atteint 0,7 (19,

20, 21). La valeur de ce coefficient dépend à la fois de l'homogénéité des items (appréciée à partir de leurs inter-corrélations) et de leur nombre (17).

Dans notre étude, le coefficient Alpha de Cronbach moyen, relatif à toutes les épreuves étudiées était de 0.84 témoignant d'une bonne homogénéité d'une façon globale.

Perspectives et Recommandations

Notre étude a permis de dégager certaines directives, visant à pallier les insuffisances de l'évaluation, et à consolider ses points forts.

- Réviser la forme et la présentation de l'épreuve
- Valider les questions
- Etudier les proportions des items selon la difficulté et la discrimination
- Adapter des épreuves au temps imparti
- Intégrer les questions au niveau des épreuves avec disciplines contributives.
- Constituer une banque d'items.

CONCLUSIONS

Notre étude a permis d'évaluer objectivement la place des disciplines dites contributives dans l'évaluation multidisciplinaire et de montrer l'intérêt de l'intégration des questions permettant une meilleure évaluation. L'analyse des questions de cette épreuve en concertation avec les enseignants permettrait une interprétation plus approfondie de ces résultats.

Conflits d'intérêts

Nous ne déclarons aucun conflit d'intérêt.

REFERENCES

1. Martin J. Aux origines de la « science des examens » (1920-1940). Histoire de l'éducation. 2002;94:177-99.
2. Valin E. La valeur des examens : Etude docimologique réalisée au Liban. Paris: UNESCO; 1961.
3. Leclercq D, Nicaise J, Demeuse M. Docimologie critique: Des difficultés de noter des copies et d'attribuer des notes aux élèves. In : Demeuse M, dir. Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation. Liège: Les Editions de l'Université de Liège; 2004. p. 273-92.
4. Gonzalez-Cabezas C, Anderson OS, Wright MC, Fontana M. Association Between Dental Student-Developed Exam Questions and Learning at Higher Cognitive Levels. J Dent

- Educ. 2015;79(11):1295-304.
5. Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH. The quality of in-house medical school examinations. Acad Med. 2002;77(2):156-61.
6. Steinert Y, Nasmith L, McLeod PJ, Conochie L. A teaching scholars program to develop leaders in medical education. Acad Med. 2003;78(2):142-9.
7. Armstrong EG, Doyle J, Bennett NL. Transformative professional development of physicians as educators: assessment of a model. Acad Med. 2003;78(7):702-8.
8. Laveault D, Grégoire J. Introduction aux théories des tests. 3^{ème} édition. Bruxelles: De Boeck Supérieur; 2014.
9. George D, Mallery P. (2003). SPSS for Windows step by step: A simple guide and reference 11.0 update. 4^{ème} édition. Boston: Allyn & Bacon; 2003.
10. George S, Haque MS, Oyebofe F. Standard setting: comparison of two methods. BMC Med Educ. 2006;6:46.
11. Caulfield M, Redden Geoffrey, Sondheimer H. Graduation Rates and Attrition Factors for U.S. Medical School Students. Anal Brief AAMC. 2014;14(5):1-2.
12. Lefebvre O. Note d'information : Réussite et échec en premier cycle [En ligne]. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, 21/11/2013 [Mise à jour le 10/04/2015 ; cité le 20/12/2015] ; [environ 7 écrans]. Disponible à l'URL : http://cache.media.enseignementsup-recherche.gouv.fr/file/2013/44/7/NI_MESR_13_10_283447.pdf
13. Arlet JB, Ranque B, Seidowski A, Passeron A, Dupeux S, Foïs E, et al. Les étudiants en médecine savent-ils répondre à des questions faciles aux examens ? Rev Med Interne. 2014;35(2):74-5.
14. Chatelain D, Charfi S, Cordonnier C, Leclercq F, Sevestre H. Étudiants en médecine et enseignement de l'anatomie pathologique : résultats d'une enquête amiénoise. ANN PATHOL. 2009;29(3):173-9.
15. Compérat EM. Mémoire Enseignement de l'Anatomie et Cytologie Pathologique en Autriche et en France. Service d'Anatomie et Cytologie Pathologique Hôpital La Pitié Salpêtrière Université Pierre et Marie Curie Paris VI [Cité le 15/12/2015] ; [environ 18 écrans]. Disponible à l'URL : <http://www.edu.upmc.fr/medecine/pedagogie/memoire/memoire%20Dr%20Compera.pdf>
16. HATIVA N. Teaching for Effective Learning in Higher Education. Dordrecht: Kluwer Academic; 2000.
17. Albanese M, Case SM. Progress testing: critical analysis and suggested practices. Adv Health Sci Educ Theory Pract. 2016;21(1):221-34.
18. Demeuse M, Henry G. L'analyse classique d'items. In : Demeuse M, dir. Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation. Liège: Les Editions de l'Université de Liège; 2004. P.173-86.
19. Peterson RA. Une méta-analyse du coefficient alpha de Cronbach. Recherche et Applications en Marketing. 1995;10(2):75-88.
20. Rodriguez MC, Maeda Y. Meta-analysis of coefficient alpha. Psychol Methods. 2006;11(3):306-22.
21. Laveault D. Soixante ans de bons et mauvais usages du alpha de Cronbach. MEE. 2012;35(2):1-7.