

**ANALYSE DOCIMOLOGIQUE DES EXAMENS EN BIOLOGIE HUMAINE
AU COURS DE SIX ANNEES ACADEMIQUES.**

**Une expérience de la Faculté des Sciences de la Santé de l'Université
d'Abomey-Calavi au Bénin.**



**GANGBO¹ Flore, NAPPORN² Clarisse, HOUNNOU¹ Gervais,
AYELO³ Gilbert**

1. Unité de Biologie Humaine, Faculté des Sciences de la Santé, Université d'Abomey-Calavi,
Cotonou, Bénin ; 01 BP 188 Cotonou

2. Département de Psychologie et des Sciences de l'Éducation

3. Ministère de la Santé, Cotonou, Bénin ; 01 BP 888 Cotonou

Correspondance : Flore GANGBO, Faculté des Sciences de la Santé, Université d'Abomey-Calavi, Cotonou,
Bénin ; 01 BP 188 Cotonou ; armandegangbo@yahoo.fr

RESUME

Introduction : L'examen écrit est l'instrument privilégié dans l'évaluation des apprentissages en Biologie Humaine à la Faculté des Sciences de la Santé (FSS) de l'Université d'Abomey Calavi au Bénin.

Objectif : L'objectif de cette étude est d'évaluer la qualité docimologique des examens écrits en Biologie Humaine à la FSS.

Cadre, matériel et méthode : Cette étude rétrospective à visée descriptive et analytique, a porté sur les notes de Biologie Humaine de six promotions successives d'étudiants de première année de médecine et de pharmacie, au cours des années académiques 06 – 07, 07 – 08 ; 08 – 09 ; 09 - 10 ; 10 – 11 et 11 – 12. Au total, 987 étudiants ont eu répondre à 149 questions réparties en 82 QCM et 67 QROC.

Nous avons calculé pour chaque épreuve d'examen et pour chacune des questions, l'indice de difficulté, l'indice de discrimination pour chaque épreuve et le degré de fiabilité de la correction.

Résultats : Au cours des six années académiques les épreuves étaient globalement de difficulté moyenne et toutes discriminantes. La fiabilité de la correction des épreuves était élevée pour : 1) l'examen entier (QCM et QROC) en 2007 – 2008 et en 2010-2011 ; 2) une partie de l'examen (QROC), en 2008 – 2009 ; 2011 – 2012

Conclusion : A partir de ces résultats, les auteurs proposent que cette expérience soit étendue aux deux sessions d'examen et à d'autres disciplines en vue de documenter d'avantage les raisons des échecs en première année et de proposer des mesures correctives adéquates.

Mots clés : évaluation des apprentissages, docimologie, fiabilité, difficulté, discrimination

SUMMARY

**DOCIMOLOGICAL ANALYSIS IN HUMAN BIOLOGY EXAMS DURING SIX YEARS. An
Experience of the Faculty of Health Sciences of Abomey-Calavi University in Benin**

Introduction: The written examination is the main instrument used in Faculty of Health Sciences (FSS) of Abomey Calavi University.

Objective: The aim of this study is to evaluate the examinations docimological quality in FSS.

Materials and methods: This study is descriptive, analytical and retrospective one, focused on medicine and pharmacy students test scores in human biology during six years : 06 – 07, 07 – 08 ; 08 – 09 ; 09 - 10 ; 10 – 11 and 11 – 12. 987 students had answered 149 questions which divided into 82 multiple choice questions and 67 briefly open questions.

For each of the test questions, we calculate difficulty index, discrimination index, rate and reliability, the difficulty index and discrimination index.

Results: Over the past six academic years: events were generally of average difficulty, they were all discriminating, reliable proofreading was high: 1) fully (for both parties and QROC QCM) in 2007 - 2008 and 2010 - 2011, 2) partially (for the QROC), 2008 - 2009 2011-2012.

Conclusion: Based to these results, the authors propose to extend this experience to the two examination sessions and other disciplines to benefit document the reasons for failure in the first year and propose corrective measures

Keywords: evaluation of learning, docimological quality, reliability, difficulty, discrimination

INTRODUCTION

A la Faculté des Sciences de la Santé, l'enseignement de la Biologie Humaine porte sur l'histologie, l'embryologie et la cytologie.

Des notes de cours comportant les objectifs, les méthodes d'apprentissage et d'évaluation sont mises à la disposition des étudiants.

L'évaluation des apprentissages est sanctionnelle et se déroule en deux sessions, l'une à la fin de l'année académique et l'autre, de rattrapage, environ quatre à six semaines plus tard.

La première session est organisée en deux temps dits « examens partiels ». Pour la 1^{ère} session, les notes des premier et deuxième partiels interviennent respectivement pour 40% et 60 %. La note de la 2^{ème} session compte pour 100%.

L'examen écrit est l'instrument prépondérant dans l'évaluation des apprentissages à la FSS. Il est composé de questions à choix multiples (QCM), de questions à réponses ouvertes et courtes (QROC), de schémas à annoter ou à réaliser.

Depuis une décennie environ, les étudiants se plaignent de leurs notes d'examen alors qu'ils ont le sentiment d'y avoir bien travaillé.

Plusieurs facteurs peuvent être incriminés dont le nombre croissant des effectifs et les qualités docimologiques de l'évaluation.

Face à ces malaises, nous avons voulu apprécier les qualités docimologiques de ces examens, en étudiant notamment leur niveau de difficulté, de discrimination, de degré de fiabilité de la correction.

- **L'indice de difficulté** correspond au pourcentage d'étudiants ayant répondu correctement à la question.
- **L'indice de discrimination** permet de distinguer les étudiants forts et des étudiants faibles [1].

- **La validité** est centrée sur la collecte partielle des données scientifiques provenant de sources multiples et se rapporte à des arguments logiques en vue d'attribuer des interprétations significatives aux données d'évaluation [2].
- **La fiabilité** se réfère à la reproductibilité des notes de l'examen. Lorsqu'elle est élevée, elle indique que si l'examen devait être répété dans le temps, les étudiants devraient avoir les mêmes que celles qu'ils ont obtenu la première fois [2].

L'objectif général de la présente étude était d'évaluer les qualités docimologiques, des examens partiels de Biologie Humaine des six promotions successives d'étudiants de première année de médecine et de pharmacie, de 2006 à 2012.

Plus spécifiquement, il s'agissait de calculer, pour les six premiers examens partiels de la première année de médecine et de pharmacie de 2006 à 2012 :

- les indices de difficulté, de discrimination de chaque épreuve et de chaque question d'examen et
- les indices permettant d'apprécier le degré de fiabilité de la correction.

CADRE, MATERIEL ET METHODE

Cette étude rétrospective à visée descriptive et analytique, s'est déroulée de Janvier 2006 à juin 2012 à l'Unité de Formation et de Recherche en Médecine de la FSS.

Elle a porté sur l'échantillonnage exhaustif des copies des examens partiels réalisés en 1^{ère} année de médecine et de pharmacie, au cours des six années académiques successives, 06 – 07, 07 – 08 ; 08 – 09 ; 09 - 10; 10 – 11 et 11 – 12.

Au total, 947 étudiants ont répondu à 149 questions réparties en 82 QCM et 67 QROC (tableau 1).

Tableau 1 : Nombre d'étudiants et épreuves d'examen

| ANNEE | 06-07 | 07-08 | 08-09 | 09-10 | 10-11 | 11-12 | Total |
|-------------------------|-------|-------|-------|-------|-----------|------------|------------|
| NOMBRE ETUDIANTS | 114 | 118 | 128 | 141 | 242 | 244 | 947 |
| EPREUVES | | | | | | | |
| - QCM | 20 | 11 | 5 | 11 | 0 | 35 | 82 |
| - QROC | 5 | 15 | 13 | 9 | 20 | 5 | 67 |
| - QCM + QROC | 25 | 26 | 18 | 20 | 20 | 40 | 149 |

Dans notre étude, l'effectif le plus élevé d'étudiants a été noté en 2011-2012. Le plus grand nombre de QCM et de QROC a été posé respectivement en 2011-2012 (n = 35) et 2009 - 2010 et 2010-2011 (n = 20).

Variables étudiées

Pour chacune des épreuves d'examen, nous avons calculé :

- les statistiques descriptives : mesures de tendances centrales et mesures de variabilité ;
- les indices de difficulté, de discrimination et de fiabilité de la correction à travers les paramètres suivants :
 - le Coefficient de Kuder – Richardson 20, lorsque l'épreuve est composée de questions à choix de réponses (QCM) ;
 - le Coefficient alpha (α) de Cronbach, lorsque l'épreuve est composée de questions ouvertes.

Pour chacune des questions d'examen, nous avons également calculé les indices de difficulté, de discrimination.

Analyse et interprétation des données

L'épreuve et la question sont classées :

- faciles si : indice de difficulté > 75 %
- de difficulté moyenne si : 25 % ≤ indice de difficulté ≤ 75 %
- difficiles si : indice de difficulté < 25 %

L'épreuve et la question sont classées :

- discriminantes si : indice de discrimination > 0,20 %
- non discriminantes si : - 1 ≤ indice de discrimination ≤ 0,20

La fiabilité de la partie QCM de l'épreuve est :

- élevée si : KR20 ≥ 0,70
- faible si : KR20 < 0,70

La fiabilité de la partie QROC de l'épreuve est dite :

- élevée si : Coefficient alpha (α) de Cronbach ≥ 0,70
- faible si : Coefficient alpha (α) de Cronbach < 0,70

Le dépouillement des copies d'examen a été manuel. Les calculs et le traitement des données ont été réalisés avec le logiciel Excel 2007. Le test de chi-2 nous a permis de comparer les données.

RESULTATS

Mesures de tendances centrales et mesures de variabilité

Tableau 2 : Mesures de tendances centrales et de variabilité

| VARIABLE | ANNEE 06-07 | 07-08 | 08-09 | 09-10 | 10-11 | 11-12 |
|----------------------|----------------|-------|--------------|------------|-------------|-------------|
| Note maximale | 15,66 | 13,92 | 18,83 | 17,5 | 19 | 17,50 |
| Note minimale | 3,33 | 1,66 | 1,39 | 0,5 | 0,5 | 1,00 |
| Moyenne | 9,04 | 8,25 | 10,80 | 7,07 | 9,7 | 9,41 |
| Mode | 10,70 | 7,94 | 9,30 | 9,64 | 10,5 | 11,00 |
| Médiane | 9,00 | 8,18 | 10,81 | 7,58 | 11,0 | 9,50 |
| Ecart type | 3,10 | 2,50 | 3,97 | 2,67 | 3,7 | 3,30 |

Au cours des six années académiques, la moyenne de la classe a atteint au moins 10 sur 20 une seule fois, en 2008 – 2009 ; la plus forte note a été obtenue en 2008 – 2009 et la plus faible en 2009 – 2010 et 2010 – 2011 (tableau 2).

Discrimination et difficulté des épreuves d'examen

L'indice de discrimination était respectivement de : 0,3 ; 0,27 ; 0,5 ; 0,5 ; 0,5 et 0,7.

L'indice de difficulté était respectivement de : 45,2 ; 41,3 ; 53,9 ; 37,3 ; 48,7 et 47.

Tableau 3 : Discrimination et difficulté des épreuves d'examen

| ANNEE | 06-07 | 07-08 | 08-09 | 09-10 | 10-11 | 11-12 | Total |
|-----------------------|--------|--------|--------|--------|--------|--------|------------|
| n = | n = 25 | n = 26 | n = 18 | n = 20 | n = 20 | n = 40 | n = 149 |
| DISCRIMINATION | | | | | | | |
| - discriminantes | 16 | 16 | 16 | 10 | 19 | 31 | 108 |
| - non discriminantes | 9 | 10 | 2 | 10 | 1 | 9 | 41 |
| DIFFICULTE | | | | | | | |
| - faciles | 3 | 2 | 2 | 1 | 0 | 3 | 11 |
| - difficulté moyenne | 20 | 15 | 16 | 10 | 16 | 30 | 107 |
| - difficiles | 2 | 9 | 0 | 9 | 4 | 7 | 31 |

La proportion des questions discriminantes est de 72,5% (108/149) et supérieure à celle des questions non discriminantes (27,5%) (Tableau 3).

La proportion des questions de difficulté moyenne est de 71,8% (107/149) et supérieure à celle des questions faciles (7,4%) et des questions difficiles (20,8%). (Tableau 3).

Niveau de difficulté des QCM et des QROC

Tableau 4a : QCM, QROC et niveau de difficulté

| | | nombre de | | |
|----------------|-------------|----------------|----------------------|----------------|
| | | ← faciles | ↓ difficulté moyenne | → difficiles |
| ANNEE | | | | |
| 06 - 07 | QCM | 3 /20 | 15 /20 | 2 /20 |
| | QROC | 0 / 5 | 5 / 5 | 0 / 5 |
| 07 - 08 | QCM | 0 /11 | 2 /11 | 9 /11 |
| | QROC | 2 /15 | 13 /15 | 0 /15 |
| 08 - 09 | QCM | 1 / 5 | 4 / 5 | 0 / 5 |
| | QROC | 1 /13 | 12 /13 | 0 /13 |
| 09 - 10 | QCM | 0 /11 | 3 /11 | 8 /11 |
| | QROC | 1 / 9 | 7 / 9 | 1 / 9 |
| 10 - 11 | QROC | 0 /20 | 16 /20 | 4 /20 |
| 11 - 12 | QCM | 2 /35 | 26 /35 | 7 /35 |
| | QROC | 1 / 5 | 4 / 5 | 0 / 5 |
| TOTAL | | 11/ 149 | 107/ 149 | 31/ 149 |

Environ ¾ des QROC et plus, étaient de difficulté moyenne. Les étudiants ont eu plus de difficultés à répondre aux QCM au cours des années 07-08 et 09 - 10 (Tableau 4a).

Niveau de discrimination des QCM et des QROC

Le plus grand nombre de QCM non discriminants a été observé au cours de l'année académique 09-10 (Tableaux 4b et 4c).

Sur les 108 questions discriminantes, 83% sont de difficulté moyenne (Tableaux 4b et 4c).

Sur les 41 questions non discriminantes, les proportions de questions difficiles et de questions faciles sont équivalentes, respectivement 43,9% (18 /41) et 41,5% (17/41) (Tableau 4b).

Tableau 4b : QCM, QROC et niveau de discrimination

| ANNEE | | Nombre de questions | |
|---------|--------------|---------------------|--------------------|
| | | discriminantes | non discriminantes |
| 06 - 07 | QCM | 13/20 | 7 /20 |
| | QROC | 3/ 5 | 2/ 5 |
| 07 - 08 | QCM | 5/11 | 6/11 |
| | QROC | 11/15 | 4/15 |
| 08 - 09 | QCM | 3/ 5 | 2/ 5 |
| | QROC | 13/13 | 0/13 |
| 09 - 10 | QCM | 2/11 | 9/11 |
| | QROC | 8/ 9 | 1/ 9 |
| 10 - 11 | QROC | 19/20 | 1/20 |
| 11 - 12 | QCM | 27/35 | 8/35 |
| | QROC | 4/ 5 | 1/ 5 |
| | TOTAL | 108/149 | 41/49 |

Niveau de difficulté des QCM et des QROC discriminantes

Tableau 4c : QCM et QROC discriminantes et niveau de difficulté

| Questions discriminantes et nombre de difficultés | | faciles | moyenne | difficiles |
|---|--------------|--------------|---------------|---------------|
| 06 - 07 | QCM | 1/ 13 | 11/ 13 | 1/ 13 |
| | QROC | 0/ 3 | 3/ 3 | 0/ 3 |
| 07 - 08 | QCM | 0/ 5 | 0/ 5 | 5/ 5 |
| | QROC | 1/ 11 | 10/ 11 | 0/ 11 |
| 08 - 09 | QCM | 0/ 3 | 3/ 3 | 0/ 3 |
| | QROC | 1/ 13 | 12/ 13 | 0/ 13 |
| 09 - 10 | QCM | 0/ 2 | 0/ 2 | 2/ 2 |
| | QROC | 1/ 8 | 7/ 8 | 0/ 8 |
| 10 - 11 | QROC | 0/ 19 | 16/ 19 | 3/ 19 |
| 11 - 12 | QCM | 1/ 27 | 24/ 27 | 2/ 27 |
| | QROC | 0/ 4 | 4/ 4 | 0/ 4 |
| | TOTAL | 5/108 | 90/108 | 13/108 |

Parmi les QCM et les QROC qui discriminent bien les étudiants forts des étudiants faibles :

- 80% au moins des QROC étaient de difficulté moyenne au cours des années 07-08 ; 08-09 ; 09 – 10 ; 10-11 et 11-12 ;
- 80% au moins des QCM étaient de difficulté moyenne au cours des années 06-07 et 11-12 (Tableau 4c) ;
- 80% au moins des QCM étaient difficiles en 2007 - 2008 (Tableau 4c).

Niveau de difficulté des QCM et des QROC non discriminantes

Parmi les QCM et les QROC qui ne permettaient pas de discriminer les étudiants forts des étudiants faibles,

- 60% au moins des QCM étaient difficiles au cours des années 07 – 08 ; 09 – 10 et 11 – 12 (Tableau 4d) ;
- 60% au moins des QCM étaient de difficulté moyenne en 2006 - 2007 (Tableau 4d).

Tableau 4d : QCM et QROC non discriminantes et niveau de difficulté

| Questions non discriminantes et nombre de difficultés | | | | |
|---|-------------|-------------|--------------|--------------|
| ANNEE | | faciles | moyenne | difficiles |
| 06 - 07 | QCM | 2/ 7 | 4/ 7 | 1/ 7 |
| | QROC | 0/ 2 | 2/ 2 | 0/ 2 |
| 07 - 08 | QCM | 0/ 6 | 2/ 6 | 4/ 6 |
| | QROC | 1/ 4 | 3/ 4 | 0/ 4 |
| 08 - 09 | QCM | 1/ 2 | 1/ 2 | 0/ 2 |
| | QROC | 0/ 0 | 0/ 0 | 0/ 0 |
| 09 - 10 | QCM | 0/ 9 | 3/ 9 | 6/ 9 |
| | QROC | 0/ 1 | 0/ 1 | 1/ 1 |
| 10 - 11 | QROC | 0/ 1 | 0/ 1 | 1/ 1 |
| 11 - 12 | QCM | 1/ 8 | 2/ 8 | 5/ 8 |
| | QROC | 1/ 1 | 0/ 1 | 0/ 1 |
| TOTAL | | 6/41 | 17/41 | 18/41 |

Fiabilité de la correction des examens

Tableau 5 : Indices de fiabilité de la correction

| ANNEE | EFFECTIF DES ETUDIANTS | QCM | QROC |
|----------------|------------------------|--------------------|---------------------|
| 06 - 07 | 114 | KR20 = 0,48 | Alpha = 0,62 |
| 07 - 08 | 118 | KR20 = 0,92 | Alpha = 0,73 |
| 08 - 09 | 128 | KR20 = -1,77 | Alpha = 0,80 |
| 09 - 10 | 141 | KR20 = 0,46 | Alpha = 0,68 |
| 10 - 11 | 242 | -- | Alpha = 0,81 |
| 11 - 12 | 244 | KR20 = 0,35 | Alpha = 1,13 |

Au cours des six années académiques, la fiabilité de la correction des épreuves était élevée pour :

- ♦ l'examen entier (QCM et QROC) en 2007 – 2008 et en 2010-2011
- ♦ une partie de l'examen (QROC), en 2008 – 2009 ; 2011 – 2012 (tableau 5).

DISCUSSION

TYPE D'EXAMEN ET NOTES OBTENUES

Dans notre étude, le nombre de questions a varié entre 18 et 40 : il s'agit le plus souvent d'une combinaison des QCM et des QROC. Plusieurs études ont été réalisées pour estimer le nombre d'évaluations nécessaires pour parvenir à un jugement fiable global [3, 4].

Pour NORCINI et coll., il faut augmenter le nombre de questions à environ soixante – dix pour obtenir un taux de fiabilité acceptable [4], ce qui du reste est difficilement réalisable, surtout si la correction est manuelle.

La longueur des énoncés pourrait être un facteur déterminant de son pouvoir discriminant : plus l'énoncé est long, moins la question aurait la chance d'être discriminante.

Dans notre étude, l'effectif de la promotion des étudiants de la première année a varié de 114 à 244 et la moyenne des notes était relativement peu élevée. Cette situation pourrait faire incriminer aussi bien les méthodes d'enseignement que les apprentissages effectués par les étudiants eux mêmes.

Le type d'examen et la longueur des questions, les notes de classe, la qualité docimologique de l'ensemble de l'épreuve d'examen ne suffisent pas pour expliquer les résultats. D'autres facteurs peuvent permettre de poursuivre la réflexion. En effet, depuis 2007, il y a eu une démocratisation des inscriptions en médecine ; ainsi toutes les séries de baccalauréat peuvent s'y inscrire. L'analyse de résultats pourrait être réalisée par rapport aux prérequis du baccalauréat.

La même étude pourrait être conduite pour les étudiants de 2^{ème} année, étendue à la même cohorte afin de rechercher d'éventuelles améliorations.

Diverses expériences pédagogiques, qui se sont déroulés à des moments différents par rapports à l'évaluation, ont montré leur influence sur l'amélioration de l'apprentissage et sur les notes des étudiants [5, 6, 7, 8].

Suite à une étude menée à la Faculté des Sciences de la Santé, nous avons montré que les étudiants entraînés à l'auto évaluation en Biologie Humaine ont répondu plus facilement aux questions que les étudiants non entraînés ; par contre qu'il y a peu de différences entre les deux groupes au niveau de la moyenne, des notes minimales et maximales

et autant de questions faciles discriminantes que des questions faciles non discriminantes (résultats non publiés).

Dans une étude portant sur l'impact de l'utilisation du CD ROM sur les notes à l'examen chez des étudiants de première année en médecine, l'étudiant pour obtenir une note supérieure ou égale à 10 sur 20, devait avoir participé à au moins six séances d'entraînement [9].

Une enquête menée auprès de 440 étudiants de première année en médecine a montré que le fait de suivre des cours préparatoires en anatomie macroscopique et en histologie leur permettait de mieux réussir ces matières, une fois qu'ils entament le cycle régulier [10]. Dans une étude menée chez des étudiants en première année de médecine, des auteurs ont prouvé que l'utilisation du microscope virtuel améliore de manière significative le rendement des élèves et l'efficacité de leur apprentissage [11].

De même, les notes du groupe expérimental de 69 étudiants de première année de médecine dentaire exposés à l'auto apprentissage assisté par ordinateur étaient significativement meilleures à celles du groupe témoin de 347 étudiants ayant suivi le cursus classique [12].

L'utilisation de la microscopie virtuelle dans l'enseignement de l'histologie dans le programme d'histologie en première année des 120 étudiants en médecine, s'est révélée très bénéfique [13].

QUALITE DOCIMOLOGIQUE DE L'ENSEMBLE DE L'EPREUVE D'EXAMEN

Dans notre étude, les épreuves d'examen ont été de difficulté moyenne et toutes discriminantes. La réussite des examens était donc à la portée de tous les étudiants.

L'effectif des étudiants est un facteur déterminant de la fiabilité de la correction; en effet, plus le nombre de copies est élevé, plus la correction est fiable et vice versa. Dans notre étude, la correction était entièrement fiable en 07 – 08 et partiellement en 08 – 09 ; 10 – 11 ; 11 – 12 avec respectivement des effectifs de 118, 128, 242 et 244.

Un coefficient de fiabilité de 0,80 ou plus est généralement considéré comme suffisamment élevé pour porter un jugement global à être utilisés dans les processus décisionnels [14].

Un taux élevé de fiabilité est lié à un niveau élevé de variance. Les facteurs de faible taux de fiabilité sont en principe :

- le faible nombre d'étudiants, lesquels risquent de former une population alors homogène
- un nombre insuffisant de questions dans l'examen
- un examen dont la facture et le mode sont tels qu'ils ne permettent pas de déceler des différences dans une population hétérogène
- lorsque l'examen mesure diverses compétences qui n'ont pas de lien entre eux.

Or, la fiabilité est un aspect important de la validité. Si le résultat final de l'évaluation (réussite ou d'échec) ne peut être reproduit à un certain niveau de certitude élevé, l'interprétation significative des scores du test est dou-

teuse et les preuves de validité est compromise [2].

Dans notre étude, les épreuves d'examen dont les corrections n'étant pas toutes fiables ne sont pas non plus valides.

Dans une population de 78 étudiants de 4^{ème} année de médecine ayant eu un examen à base de QCM, des auteurs ont obtenu au bout de 6 mois d'intervalle une bonne fiabilité inter-évaluateurs de 0,81 à 0,82 et une fiabilité test-retest modérée de 0,59 à 0,74 [15].

Par contre JANDAGHI dans son étude portant sur les bulletins examen de 364 élèves de première année du secondaire, a obtenu un coefficient de fiabilité de 0,975 et un coefficient de Kendal de 0,54, signifiant un bon degré d'accord inter juges [1].

CONCLUSION

Nos résultats ouvrent de nouvelles perspectives de recherche qui pourraient porter sur :

- du point de vue de la psychologie cognitive, les catégorisations et opérations mentales nécessaires pour s'approprier la biologie humaine ;
- les méthodes de travail des étudiants de première année
- la structure de la discipline biologie humaine (représentations préalables, déconstructions, reconstructions...)

REFERENCES BIBLIOGRAPHIQUES

- 1) Jandaghi G (2011): Assessment of Validity, Reliability and Difficulty Indices for Teacher-built Physics Exam Questions in First Year High School. *Arts and Social Sciences Journal*; 16 : 1 – 4 ; 2011
- 2) Downing SM (2003): Validity: on the meaningful interpretation of assessment data. *Medical Education* 2003; 37: 830–837
- 3) Wass V, Vleuten C, Shatzer J, Jones R (2001): Assessment of clinical competence. *Lancet*. 2001;357:945–949
- 4) Norcini JJ (2002): The death of the long case? *British Medical Journal*. 2002;324:408–409
- 5) Harris T, Leaven T, Heidger P, Kreiter C, Duncan J, Dick F (2001): Comparison of a virtual microscope laboratory to a regular microscope laboratory for teaching histology. *Anat Rec* 2001; 265 (1): 10–4.
- 6) Heidger PM Jr, Dee F, Consoer D, Leaven T, Duncan J, Kreiter (2002): Integrated approach to teaching and testing in histology with real and virtual imaging. *Anat Rec* 2002; 269 (2): 107–12.
- 7) Dee FR, Lehman JM, Consoer D, Leaven T, Cohen MB (2003): Implementation of virtual microscope slides in the annual pathobiology of cancer workshop laboratory. *Hum Pathol* 2003; 34(5):430–6
- 8) Blake CA, Lavoie HA, Millette CF (2003): Teaching medical histology at the University of South Carolina School of Medicine: transition to virtual slides and virtual microscopes. *Anat Rec* 2003; 275 (1): 196–206.
- 9) Gangbo F, Laleye A et Darboux R (2002) : Innovation pédagogique : le CD ROM dans l'enseignement de l'histologie à la Faculté des Sciences de la Santé Cotonou. *Opinions des étudiants et effets sur leurs notes d'examen. Le Bénin Médical* ; 20 : 19 - 22
- 10) Forester JP, Mc Whorter DL, Cole MS (2002): The relationship between premedical coursework in gross anatomy and histology and medical school performance in gross anatomy and histology. *Clin Anat*. 2002 Mar; 15 (2):160-4
- 11) Krippendorf BB, Lough J (2005): Complete and rapid switch from light microscopy to virtual microscopy for teaching medical histology. *Anat Rec B New Anat* 285 (1): 19-25; 2005.
- 12) Rosenberg H, Kermalli J, Freeman E, Tenenbaum H, Locker D, Cohen H (2006) : Effectiveness of an Electronic Histology Tutorial for First-Year Dental Students and Improvement in "Normalized" Test Scores. *J Dent Educ*. 70 (12) : 1339 - 45
- 13) Harry R. Goldberg and Renee Dintzis (2007): The positive impact of team-based virtual microscopy on student learning in physiology and histology. *Adv Physiol Educ* 31:261-265, 2007
- 14) Downing SM (2004): Reliability on the reproducibility of assessment data. *Medical Education*. 2004;38 : 1006 –1012
- 15) Sanju George, M Sayeed Haque, Femi Oyeboode (2006). Standard setting: Comparison of two methods. *BMC Medical Education*, 6 (46) : 1 - 6