



Université François Rabelais - Tours
UFR Arts et Sciences Humaines
Département des Sciences de l'Éducation et de la Formation

Année Universitaire 2011-2012

LE POUVOIR DISCRIMINANT DES EPREUVES ECRITES LORS D'UN EXAMEN D'ENTREE CLASSANT.

Recherche d'amélioration par une approche docimologique.

Présenté par
Philippe PELLETIER

Sous la direction de Josette LAYEC
Chargée de cours à l'université François RABELAIS

En vue de l'obtention du
Master Professionnel 2^{ème} année - Arts, Lettres & Langues
Mention Sciences de l'Éducation et de la Formation tout au long de la vie
Spécialité, Stratégies et Ingénierie en Formation d'Adultes

REMERCIEMENTS

Avant d'aborder notre propos, je tiens à remercier les personnes sans lesquelles ce travail n'aurait pas été possible :

Madame Josette LAYEC, ma directrice de mémoire, pour la pertinence de ses remarques, le respect qu'elle a accordé à ma démarche peu classique et l'autonomie qu'elle a bien voulu me laisser,

Ma compagne Laurence et mes parents, pour leur disponibilité, leur écoute, leur patience et l'amour qu'ils m'ont témoigné malgré un contexte parfois difficile,

Mes enfants, déjà grands, pour leur compréhension empathique d'un père peu disponible,

Mes amis que j'ai délaissés et qui ont respectés mon engagement dans ce parcours,

Et ma Directrice d'établissement, Madame ROUAULT, pour la liberté qu'elle m'a accordée dans le cadre de ma fonction d'enseignant formateur et son implication dans la faisabilité de cette formation.

*« Choisir ! C'est l'éclair de l'intelligence. Hésitez
vous ?... tout est dit, vous vous trompez. »*

Honoré de BALZAC né à Tours (1799 – 1850)

« L'illustre Gaudissart » (1833)

SOMMAIRE

Remerciements.....	2
Sommaire	4
Introduction.....	6
Partie 1	7
I. Un contexte et une idée de départ inscrite dans le champ conceptuel de la docimologie.	7
1. Le terrain de l'étude : un établissement de formation professionnelle.	7
2. La docimologie, une science de l'éducation et de la formation issue de la psychologie expérimentale : présentation du concept et notions associés.....	23
Partie 2	45
II. Problématisation de la question de départ : étude détaillée de l'examen et compléments conceptuels.	45
1. Le constat de départ : une question d'efficacité de la sélection.....	45
2. « Les premiers seront les derniers » ?	46
3. L'examen et les questions qui le compose : la notion de sélection et le concept associé de discrimination	54
4. Un candidat, un jury et des épreuves :Les trois composantes de l'examen d'entrée... ..	61
5. Formulation de la problématique	78
Partie 3	81
III. La recherche : une étude docimologique portant sur le pouvoir discriminant des items de l'épreuve écrite	81
1. Les buts de l'étude.	81
2. L'échantillon exploité et les méthodes utilisées	81
3. Résultats de l'étude	87
4. Discussion :les QCM et les CROC sont les plus performants.....	97
5. Critique et limites et de l'étude.....	105
6. conclusion	105
Partie 4	107
IV. Propositions d'améliorations et possibilités de réinvestissement professionnel... ..	107
1. Propositions d'amélioration de l'épreuve écrite	107
2. Réinvestissement professionnel de l'étude	112
3. Conclusion générale.....	113
Références et Index.....	115
Table des Figures	115
Références bibliographiques	116
ANNEXES	123
I. Annexe 1.....	124
II. Annexe 2.....	125

III.	Annexe 3.....	126
IV.	Annexe 4.....	129
V.	Annexe 5.....	130
VI.	Annexe 6.....	146
	Table des Matières	164

Introduction

La question des concours d'entrée dans les écoles de formations spécialisées est un problème récurrent en France depuis plusieurs dizaines d'années. Le mode de sélection appliqué à ceux-ci alimente à juste titre, une polémique sur la fiabilité de ces épreuves. En tant que formateur participant au jury de ce type d'épreuves, nous avons été interpellés et nous nous sommes intéressés aux possibilités d'améliorations à apporter à ces examens d'entrée classant.

Pour cela, nous avons choisi d'utiliser une démarche de recherche consistant en une étude docimologique rétrospective, portant sur les résultats observés lors des sept dernières années d'exercices du concours d'entrée à une école d'Infirmiers Anesthésiste. Cette méthodologie de recherche est peu courante dans le champ des sciences humaines. En effet, elle ne se situe pas dans le cadre de la recherche-action et de l'analyse d'entretiens des acteurs concernés, couramment utilisée. Nous tenons à préciser qu'il s'agit d'un choix personnel dicté par l'utilisation de données essentiellement numériques et réclamant des analyses statistiques. Une approche plus sociologique aurait été effectivement possible, mais dans ce cas il aurait été question d'une autre étude, centrée sur les acteurs, différente de celle qui nous motive, elle n'aurait pas été représentative de l'étude que nous voulons mener.

Nous allons présenter dans une première partie, le contexte de l'étude et l'émergence de la question de départ. Un détour sera fait par le rôle et la formation des Infirmiers Anesthésistes dans la Communauté Européenne pour rester fidèle à l'esprit européen de la formation qui nous a été dispensée dans ce cursus de master. Puis nous aborderons les éléments nécessaires à la compréhension de la problématique générale de la docimologie.

La seconde partie nous mènera de la question de départ vers la pose de la problématique de travail au travers d'une étude détaillée de l'examen et la précision de certains concepts.

La recherche en elle-même sera développée lors de la troisième partie, en précisant ses buts, ses résultats et la discussion que nous pouvons faire de ceux-ci.

Enfin, la quatrième partie fera état des préconisations que nous avons élaborées en termes de stratégie et d'ingénierie de formation, avant de conclure

Partie 1

I. Un contexte et une idée de départ inscrite dans le champ conceptuel de la docimologie.

1. LE TERRAIN DE L'ETUDE : UN ETABLISSEMENT DE FORMATION PROFESSIONNELLE.

1.1. L'observateur

Présenter le terrain de ce travail nous permet de préciser notre position d'observateur. Dans un premier temps je tiens à présenter mon parcours professionnel. Un baccalauréat de filière scientifique en poche, je cherche ma voie durant trois années dans des facultés de sciences naturelles et expérimentales puis en médecine. Suite à ces expériences, j'opte pour une formation d'Infirmier Diplômé d'Etat qui m'amène à obtenir un poste d'infirmier panseur dans un bloc opératoire. Rapidement, l'envie d'évoluer dans cette profession me conduit à suivre durant deux ans les cours de l'école d'Infirmiers Anesthésistes. Après quatorze ans d'exercice dans cette spécialité, durant lesquelles j'ai pu encadrer bon nombre d'étudiants se destinant à l'anesthésie, je postule au poste vacant d'enseignant-formateur à l'Ecole Régionale d'Infirmiers Anesthésistes. Ma persistance dans cette voie m'engage dans une formation de cadre de santé dont le diplôme est exigé pour cette fonction. Cela fait maintenant onze ans que j'occupe ce poste.

Le terrain auquel nous allons faire référence au cours de ce mémoire est un établissement de formation paramédicale préparant des étudiants à une spécialité infirmière. Nous occupons actuellement dans cet établissement la fonction « d'enseignant » (selon le terme en vigueur dans l'arrêté régissant le régime des études s'y rattachant¹). Le substantif de « *formateur* » nous paraît complémentaire à celui « *d'enseignant* », même si, en accord

¹ Arrêté du 17 janvier 2002 relatif à la formation conduisant au Diplôme d'Etat d'Infirmier Anesthésiste

avec Philippe Perrenoud et Donald Schön² ce vocable ne revêt pas la même signification au regard du rôle professionnalisant qui est le nôtre auprès des étudiants. En effet notre rôle de praticien réflexif nous amène à considérer l'étudiant non pas seulement comme un simple récepteur - enregistreur de connaissances théoriques, mais plutôt comme un élément complexe nécessitant plusieurs approches pédagogiques intégrant les dimensions professionnelles telles que les définit Jean-Gilles Boula³

*« la professionnalisation du savoir soignant s'ordonne sur les quatre opérations de la conscience intentionnelle que sont : la **connaissance empirique** (liée à l'expérience, aux situations professionnelles et au savoir-faire), la **connaissance intellectuelle** (la capacité de nommer ses actes propres et de manier avec pertinence les concepts propres à la profession soignante ou infirmière en termes de connaissances déclaratives et procédurales), la **connaissance rationnelle** (une attention particulièrement alertée au raisonnement professionnel et à ses conséquences), et enfin la **connaissance responsable** pour décider et agir en connaissance de cause (interrogation ou réflexion éthique sur les fins). »*

Par ailleurs, la fonction de formateur nécessite la recherche d'une posture réflexive assujettie d'une prise de recul affirmée. La remise en cause de nos pratiques est un élément essentiel de progression dans notre efficience auprès des étudiants et de la profession. Nous nous inscrivons totalement dans l'expression de cette nécessité au côté de Ph. Perrenoud qui préconise :

« On peut donc souhaiter que la pratique réflexive soit la référence des innovateurs, des formateurs, des auteurs de moyens et de méthodes d'enseignement, des cadres, et qu'on ne perde aucune occasion de la stimuler, en offrant des lieux et des ressources : séminaires d'analyse de pratiques, groupes d'échanges sur les problèmes professionnels, accompagnement de projets, supervision, aide méthodologique »⁴.

Au delà de cette vision idéaliste et communautaire du praticien réflexif, le besoin d'une prise de recul est indispensable et se pose en moteur d'une évolution. Une réflexion au sein de groupes tels que le Comité d'Entente des Formations Infirmières et des Ecoles de

² Philippe Perrenoud (2001) « Développer la pratique réflexive dans le métier d'enseignant. Professionnalisation et raison pédagogique » Paris, ESF et Donald Schön (1993) ; « Le praticien réflexif. À la recherche du savoir caché dans l'agir professionnel » Montréal, Éditions Logiques.

³ Chargé de cours en Sciences Humaines - ISIS (F-Thonon-les Bains) et Webster University - Genève

⁴ Ph. Perrenoud (2001) : « Développer la pratique réflexive dans le métier d'enseignant » Paris : ESF éditeur, p.66.

Cadres (CEFIEC), le Comité d'Entente des Ecoles d'Infirmiers Anesthésistes Diplômés d'Etat (CEEIADE), et de notre société savante professionnelle⁵, doit pouvoir s'instaurer de façon à promouvoir une dynamique nationale, d'autant moins utopique que le nombre de formateurs est restreint (moins de 70 enseignants formateurs en France). Ce peut être l'occasion de mettre en commun nos projets pédagogiques et de réfléchir à une harmonisation des pratiques pédagogiques tout en laissant à chacun la liberté de nos diversités de visions de l'anesthésie qui font la richesse d'une formation de type compagnonnique. Nous entendons par là la possibilité d'échanger et de partager des terrains de stages et des intervenants d'origines et d'écoles variées voire opposées (au sens initiatique du terme) à même de promouvoir une autoformation de type sociale (expérientielle) et existentielle (un savoir être).

Au regard de la diversité de notre activité, nous emploierons donc au cours de ce travail les termes d' « enseignant », de « formateur » ou d' « enseignant-formateur » indifféremment en fonction du contexte.

1.2.Un établissement de formation et une profession en évolution

Le terrain que nous avons retenu pour ce travail de recherche est un établissement régional de formation paramédicale. La mission principale de cette école est de former à la spécialité d'anesthésie, des Infirmiers Diplômés d'Etat (IDE) confirmés. La durée des études est de deux ans à temps plein, elle nécessite trois pré-requis obligatoires⁶ :

- Etre titulaire du Diplôme d'Etat d'Infirmier ou de Sage Femme,
- Justifier d'au moins deux ans d'expérience professionnelle à temps plein au premier janvier de l'année du concours d'entrée.
- Etre admis au concours d'entrée organisé par l'école. ce dernier élément sera d'ailleurs l'objet de notre travail d'étude.

Dans les faits, la population des candidats intégrant le cursus de formation est exclusivement composée d'infirmier(e)s, leur expérience professionnelle est en moyenne de sept ans et la moyenne d'âge est de trente ans. La répartition femmes – hommes est nettement en faveur du sexe féminin. Ce dernier constat est en revanche moins marqué que dans les Instituts de Formation en Soins Infirmiers. Nous pouvons donc déduire de cette

⁵ Dans le cas qui nous intéresse la SFAR : Société Française d'Anesthésie et Réanimation.

⁶ Arrêté du 17 janvier 2002 relatif à la formation conduisant au diplôme d'Etat d'infirmier anesthésiste

approche descriptive que nous nous situons bien dans le champ de la formation pour adultes.

Il existe actuellement vingt neuf établissements de ce type en France métropolitaine et outre-mer, tous sont rattachés à un Centre Hospitalier Universitaire (CHU). Leurs tailles sont modestes, entre quinze et soixante étudiants par année, la moyenne se situant aux environs de vingt à vingt cinq. L'école que nous considérerons ici possède un agrément pour quinze étudiants par année et son taux de remplissage est proche de cent pour cent. Le nombre relativement faible de personnes à former permet la mise en place de parcours pédagogiques personnalisés.

Les missions des écoles d'infirmiers anesthésistes définies à l'article premier de l'arrêté de formation sont les suivantes :

« former des infirmiers diplômés d'Etat ou des sages-femmes diplômées d'Etat à la polyvalence des soins infirmiers dans les domaines de l'anesthésie, de la réanimation, des urgences et de la prise en charge de la douleur, mettre en œuvre la formation préparatoire aux épreuves d'admission dans les écoles d'infirmiers anesthésistes, assurer la formation continue, promouvoir la recherche et favoriser la documentation en soins infirmiers dans les domaines précédemment cités. »

L'organisation temporelle des études au cours des vingt quatre mois de formation est régie par arrêté ministériel. Celui-ci prévoit sept cents heures de cours magistraux, travaux dirigés et travaux pratiques, réparties sur vingt semaines et soixante dix semaines de stages cliniques diversifiés au sein du CHU ou d'autres hôpitaux et cliniques. Le reste du temps est réparti en deux semaines de temps de travail personnel et cinq semaines de congés par année de formation.

Au regard des pré-requis à l'inscription à l'examen d'entrée, de la proportion et de la durée des stages cliniques, l'orientation pédagogique est clairement professionnalisante dans ce type de structure. Nous pouvons remarquer à ce sujet que nos homologues formateurs canadiens préfèrent substituer le vocable d'andragogie à celui de pédagogie lorsqu'il s'agit d'adultes. En France le terme est controversé de part sa construction étymologique qui se réfère à l'homme adulte (issu du grec « andros » - ἀνδρός-) au sens de la personne de sexe masculin et non au genre humain adulte, quelque soit son sexe.

Il est à noter que la formation délivrée dans ces écoles sera intégrée au niveau Master 2 et adossée à l'université à partir de la rentrée d'octobre 2012, dans le système Licence-

Master-Doctorat (LMD) afin de satisfaire aux accords de Bologne. A cette occasion le volume horaire réservé aux stages cliniques se verra amputé de douze semaines au profit des enseignements présentiels et du temps de travail personnel guidé. L'incidence de ce changement structurel et culturel sur la professionnalisation des futurs diplômés sera à évaluer dans un avenir proche, de l'ordre de deux à trois ans. Cette évaluation devra tenir compte de l'évolution professionnelle dans le cadre des pratiques avancées prévue par la Loi Hôpital Patient Santé Territoire de 2009 (Loi « HPST »⁷) et dans lequel nous nous situons de fait, depuis de nombreuses années.

A cet état de fait s'ajoutera un déficit démographique que certains estiment à près de cinquante pour cent à l'horizon 2015, propre à élargir le champ d'activité de ces professionnels par manque de moyens humains. Cette tendance s'observe tant chez les infirmiers que chez les médecins de cette spécialité. La capacité des écoles et des universités à former des professionnels n'est pas à même de combler le retard dans le délai imparti. A titre d'exemple, la capacité totale théorique des écoles françaises est d'environ sept cents diplômés chaque année. Dans les faits, le nombre maximum atteint plutôt six cents. Si l'on considère qu'environ trois mille cinq cents professionnels vont disparaître d'ici 2015, le déficit en formation serait de l'ordre de mille cinq cents à mille sept cents personnes, soit un peu plus du quart de la population actuelle.

Le corolaire de ce constat, du point de vue du formateur, sera de choisir avec soin (de sélectionner) les postulants à intégrer les écoles susceptibles de suivre une formation dont le niveau d'exigence ne fait que s'élever.

1.3. Une évaluation en devenir

Le système actuel d'évaluation revêt cinq aspects distincts :

- Une évaluation continue des enseignements théoriques répartie en six séquences sur les deux ans d'étude. Ces épreuves sont réalisées en conditions d'examen « sur table », elles comprennent une épreuve de questions d'ordre théoriques et une question rédactionnelle de synthèse de deux heures chacune. Les corrections sont effectuées sous couvert de l'anonymat. L'obtention d'une note au moins égale à 10 sur 20 conditionne la poursuite du cursus dans la séquence suivante. La somme de ces notes participe à hauteur de vingt cinq pour cent dans le décompte des points attribués pour l'obtention du diplôme.

⁷ LOI no2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires : article n° 51

- Une évaluation clinique à l'occasion de chaque fin de stage. Dans ce cas aucune note n'est attribuée, le stage est validé ou non, assorti d'un commentaire. En cas d'échec, il peut être proposé à l'étudiant une nouvelle évaluation à distance ou bien l'exclusion de la formation.
- Une évaluation clinique notée sous la forme de deux mises en situations professionnelles (une par an) participant à hauteur de vingt cinq pour cent dans le décompte des points attribués pour l'obtention du diplôme.
- Un travail d'intérêt professionnel rédigé sous forme de mémoire (terme qui n'est pas encore employé) consistant en une étude d'une problématique professionnelle précise assujettie d'une bibliographie actualisée. Les critères d'évaluation retenus par notre établissement portent essentiellement sur la pertinence du sujet traité, la méthodologie employée et la discussion des résultats. Une présentation et une argumentation orale publique complètent le travail. Une appréciation écrite du jury sanctionne l'épreuve. Aucune note n'est attribuée mais la réalisation du travail est obligatoire pour se présenter aux épreuves du Diplôme d'Etat.
- Une épreuve finale de Diplôme d'Etat comprenant, une mise en situation professionnelle et deux épreuves écrites similaires à celles validant les séquences portant sur l'intégralité du programme de formation. Cette partie participe à hauteur de cinquante pour cent dans le décompte des points attribués pour l'obtention du diplôme.

Dans un avenir proche, la procédure d'évaluation va changer. Tout d'abord les évaluations théoriques valideront des unités d'enseignements et pourront, en cas d'échec, faire l'objet d'épreuves de rattrapage durant plusieurs années sous réserve d'un minimum requis pour le passage en troisième semestre. L'évaluation clinique sera orientée vers une évaluation des compétences ce qui nécessitera un effort de communication très important à destination des équipes soignantes. Enfin, au travail d'intérêt professionnel sera substitué un mémoire de fin d'études qui validera l'unité d'enseignement recherche.

1.4.L'Infirmier Anesthésiste et sa formation dans le Contexte Européen

Au sein de l'Union Européenne (UE), les Infirmier(e)s Anesthésistes Diplômé(e)s d'Etat (IADE) occupent une place mineure au regard de leur nombre ramené à celui des infirmiers non spécialisés. Pour exemple en France, un des deux pays Européens les mieux

dotés en Infirmiers Anesthésistes, on recense entre 6000 et 8000⁸ IADE pour environ 485000 IDE.

Par ailleurs, dans les différents pays de l'Union Européenne, il existe une grande disparité de formations et de reconnaissances professionnelles des Infirmiers Anesthésistes. La question qui nous préoccupe dans le cadre de la construction européenne, en termes de formation, est de savoir quelle est la place des écoles d'Infirmiers Anesthésistes françaises au sein de l'Europe et de leurs possibilités d'échanger au sein de l'Union.

1.4.1. La formation des IADE dans le cadre des accords de Bologne

Dans les chapitres précédents nous avons pu aborder le déroulement de la formation des IADE en France. Si l'on prend en compte la dimension Européenne des diplômes dans le cadre du LMD, qu'en est-il des correspondances du diplôme et des fonctions d'un IADE au sein de la communauté Européenne ?

Les « accords de Bologne » imposent à toute formation post-bac d'être inscrite dans le système « Licence, Master, Doctorat ». Suite à l'intégration du Diplôme d'Etat d'Infirmier français au niveau licence, capitalisant 180 ECTS (système européen de crédits transférables)⁹, les formations d'infirmiers spécialisés se devaient de suivre le même chemin. Des accords sont en cours de finalisation au niveau des ministères de la Santé et de l'Enseignement Supérieur et de la Recherche afin que les Infirmiers Anesthésistes, obtiennent une reconnaissance au niveau Master. En effet, la formation d'anesthésie s'effectue sur deux années à plein temps et sera créditée des 120 ECTS nécessaires à l'obtention du Master.

Si au niveau français, les choses semblent s'éclaircir, il en va tout autrement au niveau européen. Nous pouvons en effet remarquer qu'il existe un écart très important entre les membres de l'Union Européenne quant à la perception du rôle de l'IADE au sein des systèmes de santé, et également certaines aux réticences du corps médical à reconnaître l'importance, voire même l'existence de cette profession¹⁰.

⁸ Statistiques DREES (Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques) ou FHF (Fédération Hospitalière de France)

⁹ European Credits Transfer System

¹⁰ En Belgique par exemple, le titre d'infirmier anesthésiste existe mais n'a aucune reconnaissance statutaire professionnelle dans le déroulement des actes anesthésiques.

1.4.2. Les IADE en Europe

Si, au niveau européen, le cadre légal du diplôme infirmier est clairement établi et permet la libre circulation au sein de la communauté, en revanche celui des infirmiers spécialisés en anesthésie est beaucoup moins clair. Il est important de considérer que celui-ci :

« ... relève des directives dites du système général. Les ressortissants de l'Union Européenne, titulaire d'un diplôme d'infirmier non conforme aux obligations fixées par les directives européennes peuvent bénéficier d'une autorisation d'exercer la profession d'infirmier délivrée par le préfet de région. En cas de différences substantielles de formation, l'intéressé doit se soumettre à une épreuve d'aptitude ou d'accomplir un stage d'adaptation »¹¹

Cet extrait des directives européennes donne la mesure de la difficulté à exercer une spécialité infirmière en dehors des frontières propres à un état de l'union. Cela signifie, d'un point de vue pratique, qu'il sera nécessaire de prendre contact avec les autorités européennes compétentes en la matière, dans l'état où un IADE voudra travailler, afin d'obtenir la reconnaissance de son diplôme. Ces contraintes administratives semblent d'ores et déjà très compliquées.

1.4.3. La diversité des rôles et des formations des IADE en Europe

La diversité des formations des IADE dans l'UE est une réalité à prendre en considération. Peu de travaux rendent compte de façon précise des différences existantes. Pascal ROD¹² en 2006 présentait un tableau assez complet du sujet : seuls les Pays Bas, et le Luxembourg présentaient une durée de formation équivalente ou supérieure à celle de la France. En ce qui concerne la formation théorique, les Pays Bas et la Suède proposent un volume horaire égal ou supérieur à celui de la France.

La position de la Suède est assez particulière puisqu'elle propose une formation théorique très importante (environ 1100 heures) au regard de sa durée de formation qui est de douze mois (24 mois dont 700 heures de cours en France). Cette situation laisse à penser que les enseignements pratiques, sur le terrain, ne sont pas intégrés au cursus de formation. nous n'avons malheureusement pas pu obtenir plus de renseignements sur la formation suédoise, en particulier si ces douze mois étaient consécutifs ou bien répartis sur plusieurs années

¹¹ Directives 89/48/EEC et 92/151/EEC

¹² ROD P. (2006) « Les infirmiers anesthésistes et la réglementation européenne de la mobilité professionnelle » in Oxyimag, n° 89 août 2006

parallèlement à un emploi d'infirmier anesthésiste stagiaire. Le Luxembourg et les Pays Bas quant-à eux, proposent des formations comparables aux écoles françaises, mais le volume d'infirmiers anesthésistes formés ne représente que peu de personnels.

Il est à noter qu'au Pays Bas, à l'instar de ce qui se passe au Royaume Unis, il existe un autre type de professionnels d'anesthésie non médecins ; il s'agit des « *medwerkers* » et des « *Anaesthesia Physician Assistants*¹³ » pour lesquels une formation infirmière préalable n'est pas requise.

Les situations de l'Allemagne, du Danemark et de l'Autriche sont différentes. L'Allemagne forme des IADE de façon assez complète, mais chaque « land » applique une réglementation particulière. Le Danemark propose une formation plus courte mais la reconnaissance professionnelle est comparable à celle de la France. L'Autriche dispose d'une formation courte de 4 mois, mais combinée avec les soins intensifs¹⁴.

Les autres pays de l'UE ne forment pas de façon clairement définie des Infirmiers Anesthésistes. Dans ces cas, les IADE sont, soit inexistants, soit non différenciés des infirmiers en soins généraux, soit identifiés dans des emplois habituellement réservés à des diplômes de niveau 5 en France.

Dans tous les cas, nous n'avons pu obtenir de précisions concernant les conditions d'admissions dans les établissements de formations. Il n'est jamais fait mention de pré requis professionnels précis, ni de concours ou d'examens d'entrée. Ceci place la France dans une condition particulière, de part sa culture très ancrée d'une sélection préalable aux formations de toutes sortes.

Nous pouvons noter qu'il existe depuis 1989 une fédération internationale¹⁵ des infirmiers anesthésistes regroupant maintenant 36 pays membres dont 14 membres de l'UE. Parmi ceux-ci, moins de la moitié proposent des formations reconnues au niveau national.

1.5.Un projet européen pour les étudiants infirmiers anesthésistes ?

La grande disparité des pratiques et représentations en Europe concernant notre profession ainsi qu'une démographie particulièrement défavorable pour les IADE comme pour les

¹³ Meesen V. et Col. (2010), "Composition of the anesthesia team: a european survey" in Eur J Anesthesiol, Sept; 27(9): 773-9

¹⁴ Source Syndicat National des Infirmiers Anesthésistes (Web 2011)

¹⁵ IFNA : International Federation of Nurses Anesthesits

Médecins Anesthésistes Réanimateurs (MAR) partout dans l'UE, va nécessiter une uniformisation des pratiques et des formations :

« L'appel à l'uniformisation des pratiques au sein de l'UE est également discutée. Et si l'on se souvient qu'entre la Turquie et l'Angleterre et la France ou la Suède il existe de très fortes différences de pratiques ... il faut être vigilant à ce que le nivellement ne se fasse pas par le bas mais vers le haut ... »¹⁶

Si la Turquie ne fait pas encore partie de l'UE, l'exemple pris avec d'autres pays tels que la Roumanie ou la Slovaquie reste valable. Cette pénurie largement annoncée et l'uniformisation inquiètent effectivement beaucoup toute la profession en Europe.

S'il reste incontournable d'attendre une réglementation européenne, la possibilité d'un projet européen concernant la mobilité dans la formation des infirmiers anesthésistes est envisageable mais présente malgré tout quelques difficultés :

- La première d'entre elles concerne les pays partenaires d'une telle initiative. La situation décrite au chapitre précédent identifie à notre sens cinq états de l'Union : La Suède, le Danemark, les Pays Bas, le Luxembourg, l'Allemagne dans certains « Land » non encore identifiés et l'Autriche dans une moindre mesure. L'Angleterre pourrait également faire partie de ce groupe, mais il faudrait correctement identifier les établissements proposant un accueil en adéquation avec le rôle propre des « Anaesthesia Physician Assistants » souvent différent de celui des IADE français en particulier.
- La seconde concerne un point réglementaire français qui définit clairement le statut de « profession réglementée », interdisant la pratique de l'anesthésie à toute personne non titulaire du Diplôme d'Etat IADE. Les étudiants IADE français sont autorisés à participer aux anesthésies dans le cadre de leurs stages, mais qu'en est-il des étudiants européens non inscrits dans les écoles de l'hexagone ?
- La troisième pose la question du type de programme européen de mobilité dans lequel pourrait s'inscrire un tel projet : Programme « Léonardo Da Vinci » ou « Erasmus » ? Nous n'avons malheureusement pas de précisions à apporter quant aux reconnaissances dans le système LMD des diplômes d'Infirmiers Anesthésistes hors de l'état français. Il semblerait que la France soit le premier état à reconnaître cette formation au niveau Master. Nous restons malgré tout optimistes, le futur

¹⁶ Meesen V. et Col. (2010), "Composition of the anesthesia team: a european survey" in Eur J Anesthesiol, Sept; 27(9): 773-9

programme « Erasmus pour tous », prévu pour 2013 – 2014 dans le cadre de la « stratégie Europe 2020 » devrait permettre de ne plus faire vraiment de différence entre les sept programmes actuellement proposés.

Reste le problème de la « mobilité qualifiante » qui permettrait l'obtention des ECTS correspondants aux objectifs pédagogiques d'une telle mobilité.

- La quatrième difficulté est d'ordre purement pratique. En qualité d'enseignant formateur, nous sommes en charge d'organiser les périodes de stages pratiques, l'enseignement clinique. Le nombre de places disponibles dans les blocs opératoires agréés par l'école n'est pas très important, et juste suffisant pour nos étudiants. Dans certaines spécialités comme la chirurgie pédiatrique, le problème est aigu pour tous les établissements et donne lieu à de nombreuses négociations entre écoles et établissements hospitaliers. Cette dernière difficulté pourrait se trouver commuée en avantage si des échanges étaient réalisés entre membres de l'UE disposant de centres de chirurgie pédiatrique pouvant faire l'objet d'un agrément.

Au regard de ces différents arguments, la tâche nous paraît ardue, mais non insurmontable :

Dans un premier temps, il nous semblerait judicieux de délimiter le cadre du partenariat¹⁷ aux pays suivants : La Suède, le Danemark, les Pays Bas, le Luxembourg, et l'Allemagne dans certains « Land ». Ces états représentent en effet, avec la France, un groupe assez homogène de pratiques cliniques et de formations.

Dans un second temps, nous pensons qu'il est souhaitable d'attendre la mise en œuvre du nouveau programme « Erasmus pour tous » qui simplifiera, ou nivellera, la question des différences de reconnaissances des diplômes au sein du système LMD.

L'ingénierie de la certification et de l'évaluation ne semble pas avoir été abordée jusqu'ici et devra se dérouler sur un mode de gré à gré entre les organismes de formations et leurs instances tutélaires de certification.

Enfin, la possibilité de proposer des stages cliniques et/ou des enseignements théoriques spécifiques aux pays receveurs nous paraîtrait tout à fait enrichissante sur le plan culturel et professionnel. Nous croyons fermement que de tels échanges favoriseraient l'élaboration d'un cadre européen permettant d'uniformiser les approches du métier d'Infirmier

¹⁷ Intervention de Chantal Marquois au cours de l'UE 10

Anesthésiste dans les frontières de l'Union Européenne. Ce dernier point risque de devenir crucial dans un avenir proche afin de garantir une prise en charge anesthésique de qualité en Europe.

La mobilité européenne des étudiants est une réalité pour de nombreux cursus universitaires. La formation des étudiants Infirmiers Anesthésistes, prochainement adossée à l'Université dans le cadre de la reconnaissance dans le système LMD au niveau Master, devrait permettre cette mobilité. Les difficultés spécifiques à la profession, essentiellement liées à la diversité des pratiques, seront certainement amoindries lorsqu'un cadre réglementaire européen sera proposé. Il s'agit bien sûr d'un projet à long terme, nécessitant la rédaction de référentiels communs (activités, compétences et formations). Le brassage des diverses pratiques européennes devrait favoriser la connaissance et la reconnaissance des différences et apporter sa contribution à l'édification de tels documents.

Dans ce chapitre, le cas des professionnels diplômés a été peu abordé. Leur mobilité est théoriquement possible mais soumise à un cadre réglementaire contraignant. Les exemples courant d'expatriations temporaires en Suisse (non membre de l'Union Européenne) sont souvent bien vécues par les deux parties. Cet exemple nous permet d'envisager avec optimisme des possibilités de formations professionnelles continues dans les pays adhérents à l'Union Européenne.

1.6.L'examen à l'entrée à l'origine de la question de départ

Comme c'est souvent le cas en France pour bon nombre de formations, les candidats à suivre la formation d'Infirmiers Anesthésistes doivent « *subir avec succès les épreuves d'admission* ¹⁸ » organisées par l'établissement formateur. Ces termes sont suffisamment explicites pour exprimer les exigences des établissements de formation. La sélection, puisqu'il faut l'appeler par son nom, nécessite une clarification quant au terme utilisé. Le plus souvent, les vocables de concours d'entrée, d'examen d'entrée ou d'examens classants sont les plus utilisés. Lequel est le plus apte à définir la réalité ?

¹⁸ Expression extraite du texte de l'arrêté de formation.

1.6.1. Les épreuves d'admission : examens, concours d'entrée ou examens classants ?

Dès 1952, P. DAGUE¹⁹ s'interroge sur la cette terminologie et pose la question de la définition des examens et des concours.

Pour cet auteur, l'examen aurait pour but de vérifier si les objectifs de la formation ont été atteints et si oui, à quel niveau d'excellence, tant sur le plan des acquisitions théoriques que des compétences : *« l'art de sen servir »*. A ce titre, nous pouvons remarquer une définition de la compétence très proche de ce qu'a pu produire G. Le BOTERF trente années plus tard. Ceci défini une épreuve diagnostique quant aux acquisitions de l'étudiant lors de son cursus.

Le concours, toujours selon P. DAGUE²⁰ :

« viserait, au contraire, à sélectionner les plus aptes pour certaines tâches intellectuelles à venir. Il fournirait donc un pronostic de réussite et, comme tel serait tourné vers l'avenir. » mais... *« les concours ... font essentiellement appel aux connaissances scolaires... »*

En accord avec l'auteur, cette opposition semble d'ores et déjà inadaptée dans le sens où la somme de connaissances à produire lors de ce type d'épreuve fait plus appel à la capacité mnésique du candidat qu'à sa véritable capacité à faire preuve d'efficacité professionnelle. Il nous faut néanmoins remarquer la recherche des capacités de projection du candidat à suivre une formation dans l'avenir. Par ailleurs, la notion d'examen est assujettie à la condition de répondre à une note au moins égale à la moyenne théorique de la note de l'épreuve ou bien à une note minimale préalablement convenue.

Le concours, quant à lui, ne se réfère pas à une note minimale d'admission mais simplement au classement final, quelque soit le niveau de la dite note. Le nombre de places disponibles dans les établissements concernés peut rapidement tendre à transformer un examen en concours pour peu que la réglementation soit suffisamment imprécise.

Certains exemples d'examens d'entrée tendent à concilier les deux aspects précédemment évoqués. Ceux-ci considèrent d'une part un classement par ordre décroissant des notes, et d'autre part leur répartition autour de la moyenne des notes attribuées. Une sélection est ensuite effectuée, retenant les personnes dont la note est supérieure à la moyenne des notes obtenue à l'épreuve. Nous remarquerons qu'il s'agit là d'un système extrêmement sélectif.

¹⁹ Dague P. Le régime des examens et des concours en France : problèmes de docimologie. In: Enfance. Numéro Spécial, 1952.-pp. 449-450.

²⁰ Ibid.

En effet, si tous les candidats ont obtenu des notes régulièrement réparties et comprises entre seize et dix huit sur vingt, tout candidat dont la prestation n'a pas été évaluée à au moins dix sept sur vingt se verra recalé. Il s'agit là de cas extrêmes cités pour exemples.

Dans le cas qui nous intéresse, la réglementation nous impose de classer les candidats selon leur rang d'admission, à la condition qu'ils aient pu obtenir une note au moins égale à la moyenne pour chacune des épreuves. La conséquence de ces deux obligations nous amène à proposer le terme d'examen classant, déjà utilisé dans d'autres filières de formation, en particulier celle des études de médecine.

1.6.2. Les modalités de l'examen classant à l'entrée

Deux épreuves composent l'examen d'entrée à la formation d'Infirmiers Anesthésistes Diplômés d'Etat : Une épreuve écrite et une épreuve orale.

L'épreuve écrite

L'épreuve écrite consiste à répondre à vingt questions destinées à tester le niveau de connaissances de base du candidat. Celles-ci sont d'ordre théorique, issues de la totalité du programme de formation d'Infirmier Diplômé d'Etat. La durée de composition est d'une heure et trente minutes. Une double correction est réalisée²¹ par un médecin et un IADE. Selon un barème et une grille de correction établie en réunion préliminaire. Les items composant l'épreuve sont proposés par les membres du jury lors de cette réunion de préparation. Une note globale de la copie au moins égale à vingt sur quarante est exigée pour accéder à l'épreuve orale. En regard de ce premier niveau de sélection, cette partie écrite est appelée épreuve d'admissibilité.

L'épreuve orale

L'épreuve orale s'organise autour d'une question d'ordre professionnel choisie dans les mêmes conditions que les questions écrites. La durée est de vingt minutes, réparties en dix minutes de présentation du cas clinique et de réponses aux questions posées ; les dix minutes restantes sont réservées aux questions que le jury désire poser au candidat. Le candidat dispose d'une période de préparation de vingt minutes immédiatement avant leur prestation orale devant le jury. L'épreuve orale vise à apprécier les qualités professionnelles du candidat quant à ses aptitudes à suivre la formation avec succès. Une note minimum de vingt sur quarante est exigée pour l'admission.

²¹ Considérée comme illusoire depuis les travaux de LAIGIER et WIENBERG en 1936 (voir plus loin dans le chapitre consacré à la docimologie)

Le classement

Vient ensuite le classement par ordre de mérite. Dans un premier temps, les notes des deux épreuves sont additionnées, donnant une note sur quatre vingt. Le texte réglementaire précise qu'une note au moins égale à quarante sur quatre vingt est exigée pour déclarer le candidat admis²². Les candidats sont ensuite classés par ordre de résultats décroissants. Les quinze premiers seront déclarés admis sur la liste principale, les suivants seront déclarés admis sur la liste complémentaire. Les candidats admis sur cette dernière liste ont la possibilité d'accéder à la liste principale en cas de désistement d'un des quinze premiers reçus. Ces cas ne sont pas rares, puisque nous avons pu appeler le dernier de la liste complémentaire pour suivre la formation. La raison de ce « jeu de pousse-pousse » est essentiellement à mettre au crédit, si nous osons le dire, du manque de crédits de formation alloués aux établissements de santé par leurs différents organismes collecteurs agréés.

Le jury

Le jury est proposé par la direction de l'école. Il est composé de huit personnes réparties à parts égales entre médecins anesthésistes réanimateurs et paramédicaux spécialisés en anesthésie-réanimation. Le corps médical comprend le Directeur scientifique de l'établissement, celui-ci est obligatoirement Professeur des Universités spécialisé en anesthésie et réanimation. Les autres membres sont des Maîtres de Conférences Universitaires ou des Praticiens Hospitaliers Anesthésistes-Réanimateurs. Le corps paramédical est représenté par le Directeur de l'école, celui-ci doit être Directeur des Soins et titulaire du Diplôme d'Etat d'Infirmier Anesthésiste. Lui sont adjoints le formateur permanent de l'établissement Cadre de Santé IADE, des Cadres de Santé IADE et/ou des IADE travaillant dans les blocs opératoires, en salle de surveillance post-interventionnelle (SSPI : anciennement salle de réveil) ou en réanimation. La particularité de l'établissement qui nous intéresse est de présenter une double parité : parité médecins / paramédicaux qui est une obligation réglementaire, précisée dans l'arrêté relatif à la formation, ainsi qu'une parité femmes / hommes qui est un choix éthique de la direction de l'école.

1.7.L'origine de la question de départ

A la fin de chaque session annuelle d'examen d'entrée, une réunion de l'ensemble du jury permet de faire le bilan de l'ensemble des épreuves. A cette occasion chacun peut exprimer

²² Si l'on considère les conditions de la notation à l'épreuve orale, nous pouvons remarquer qu'un candidat très brillant à l'épreuve écrite et ayant une note inférieure à vingt à l'épreuve orale sera recalé, même si la somme des deux notes est supérieure à quarante.

son avis et ses interrogations à posteriori quant au contenu des questions proposées et la forme présentée. Lors d'un de ces récents bilans, le Directeur Scientifique s'est interrogé sur la répartition des notes de l'épreuve écrite autour de la moyenne. L'ordre du jour étant chargé, la décision fut prise que nous nous chargerions de présenter un rapport sur le sujet à l'occasion de la prochaine réunion avec nos autorités de tutelles. Comme nous pouvions nous y attendre²³, la distribution des notes tendait vers une répartition gaussienne²⁴, centrée sur une valeur légèrement supérieure à la moyenne mais néanmoins très proche de celle-ci. Gilbert. DE LANDSHEERE confirme en 1976²⁵ la normalité de ces résultats en la justifiant ainsi :

« Les résultats scolaires ou ceux des tests devant logiquement être parallèles aux potentialités des sujets, on a conclu que la répartition (...) des notes se faisait aussi selon une courbe normale »

Si dans pour une mesure physique c'est le résultat recherché, dans le cas qui nous intéresse, il est surtout le reflet d'une indécision du (des) correcteur(s) ou bien d'une ingénierie de construction de l'examen ne permettant pas de faire clairement la nuance entre les bons candidats et les « moins bons ».

C'est donc là une tendance naturelle à laquelle il faut essayer d'échapper. En effet dans la logique d'un examen classant, la recherche d'une sélection de deux populations, les bons et les moins bons, devrait conduire à une distribution des notes de type binomiale attestant de la présence de deux groupes distincts parmi les candidats.

Cette réflexion nous a conduits à exprimer la question de départ de ce travail de recherche :

"Comment peut-on améliorer la sélection des candidats à suivre une formation lors d'un examen classant ou d'un concours ?"

Avant de présenter la démarche de problématisation, quelques éclaircissements s'imposent. La question se situant clairement dans le champ de la docimologie, un détour vers les

²³ LAUGIER et WIENBERG (1927) « les facteurs subjectifs dans les notes d'examen » in Année psychologique, XXVIII pp. 236, 244.

²⁴ Loi statistique de Distribution Normale de GAUSS et LAPLACE. Cette loi démontre que pour tout phénomène naturel : plus l'échantillon est important, plus la valeur mesurée tend vers une tendance centrale. La représentation graphique est la courbe dite « en cloche » bien connue des chercheurs.

²⁵ De LANDSHEERE, (1976) « Introduction à la recherche en éducation » Ed Armand COLLIN, PARIS, 4^{ème} Ed.

fondamentaux de cette science nous paraît indispensable à la bonne compréhension de la démarche de recherche.

2. LA DOCIMOLOGIE, UNE SCIENCE DE L'EDUCATION ET DE LA FORMATION ISSUE DE LA PSYCHOLOGIE EXPERIMENTALE : PRESENTATION DU CONCEPT ET NOTIONS ASSOCIEES.

Traiter d'un sujet de docimologie nécessite de préciser le champ de cette discipline dont la représentation est souvent restrictive.

2.1.Science de la notation, oui, mais pas seulement.

Bien souvent, la docimologie est présentée comme la science des notes et de la notation. Comme nous allons le voir dans les chapitres suivants, les faits marquants de l'histoire de cette science assez récente peuvent expliquer cette représentation fréquente. Limiter le champ de la docimologie à la notation est en fait très restrictif et ne reflète pas la diversité des domaines qu'elle explore. Cette méconnaissance du grand public reflète assez fidèlement la place qui lui est réservée en France, son pays de naissance. A titre d'exemple, pour un étudiant qui chercherait à approfondir sa culture sur le sujet, la bibliothèque de l'Université François RABELAIS, par ailleurs particulièrement bien fournie, ne propose que quatre ouvrages, deux thèses et un article en réponse à la recherche du mot clé « docimologie » effectuée sur la base de données de ce service. Nous pouvons également remarquer les dates de parution de ces documents : toutes inférieures à 1972 pour les livres, et indisponibles en accès direct.

2.2.Phase une : Naissance de la docimologie, le temps des constats

Il est communément admis²⁶ que la paternité du terme « docimologie » est à mettre au crédit d'Henri PIERON²⁷. Le mot apparaîtrait pour la première en fois en 1922 (ou 1929 pour certains auteurs²⁸), construit à partir du grec « Dokimès » (examen, épreuve) et

²⁶ Martin J. (2002) : « Aux origines de la « science des examens » (1920-1940) », Histoire de l'éducation n°94

²⁷ (1881 1964)Titulaire de la chaire de physiologie des sensations au Collège de France de 1923 à 1951

²⁸ Leclercq D., Nicaise J., Demeuse M. (2004) : « Docimologie critique :Des difficultés à noter les copies et d'attribuer des notes aux élèves »p. 274

« Logos » (science). Cette « science des examens » ne sera définie qu'en 1951 dans son ouvrage « vocabulaire de la psychologie » en ces termes : « *l'étude systématique des examens (modes de notation, variabilité interindividuelle et intra-individuelle des examinateurs, facteurs subjectifs, etc.)* ». Inspirés par les travaux réalisés à la fin du XIX^{ème} siècle par Alfred BINET (sur le surmenage intellectuel et la fatigue scolaire) et Edouard TOULOUSE, dont il deviendra un de ses élèves, Henri PIERON travaille de concert avec son collègue Henri LAUGIER depuis 1922 sur l'observation des performances scolaires d'élèves testés par leurs soins. Le résultat de ces tests était comparé à la réussite au Certificat d'Etude Primaire. En comparant les deux résultats, les auteurs constatèrent qu'il n'existait aucune corrélation valable. Cette mise en évidence d'une possible inadaptation des examens à leurs objectifs initiaux remettait en cause tout un système institutionnel de notation en place depuis l'époque napoléonienne et la création du baccalauréat²⁹. A ce propos, nous pouvons faire remarquer que l'association des deux auteurs au cours de ces études, a pu alimenter la polémique de la paternité de la docimologie à H. LAUGIER plutôt qu'à H. PIERON. Néanmoins, nous avons vu précédemment que le terme de docimologie a été effectivement défini par H. PIERON.

Nous pouvons noter que ces psychologues étaient issus de la nouvelle psychologie expérimentale, précurseurs de ce qui sera plus tard nommé psychométrie. Cet élément historique explique certaines analogies avec l'édu-métrie, domaine actuel très actif de la docimologie.

Les véritables lettres de noblesse de cette nouvelle science seront écrites en 1936 avec la publication par H. LAUGIER et D. WEINBERG des résultats de « l'Enquête CARNEGIE ». Malgré une faible diffusion, l'ouvrage provoqua selon H. PIERON « *une certaine émotion*³⁰ ». Le sujet restera longtemps sensible (et l'est toujours actuellement pour certains professionnels de l'enseignement³¹) à tel point qu'en 1963 H. PIERON écrit en préambule à la présentation de l'enquête dans son ouvrage de référence³² sur le sujet :

« *Il convient d'ajouter qu'il ne s'agit point d'incriminer des hommes, mais uniquement des méthodes, et qu'aucun nouveau système ne saurait être recommandé a priori, sans avoir été expérimenté, comme cela se fait pour les méthodes thérapeutiques.* »

²⁹ 1808

³⁰ Piéron H. (1963). « *Examens et Docimologie* » : Paris : Ed. PUF ; Col. Le psychologue. P.16

³¹ Dauvisis MC. (1999) : « *Valeurs latentes des pratiques d'évaluation certificatives* » in Spirales-Revue de Recherches en Education, 1999 n°23, p7

³² Ibid.

La notion d'expérimentation évoquée dans cet extrait d' « *Examens et docimologie* » place la docimologie dans le champ de la recherche expérimentale chère à Claude BERNARD, où se plaçaient déjà ses prédécesseurs. La première phase du développement du mouvement docimologique est donc essentiellement critique³³.

« ... dès lors la docimologie s'autonomise comme discipline scientifique à part entière sur le modèle des sciences expérimentales ³⁴ »

C'est ce que nous allons pouvoir constater dans le chapitre qui suit.

2.2.1. Les constats : L'Enquête CARNEGIE, un pavé dans la mare.

En 1931, l'Institute of International Education de l'université de Columbia demande une étude sur « *Les conception, les méthodes, la technique et la portée pédagogique et sociale des examens et concours* ». La CARNEGIE Corporation est sollicitée pour le financement. Par la suite, ce travail international sera connu sous le nom « d'Enquête CARNEGIE ». Il est à noter qu'une partie des fonds utilisés pour cette enquête proviendrait de la fondation H. Ford, ce qui resituerait ce travail dans le cadre des grandes entreprises du début du XX^{ème} siècle cherchant à développer une culture d'entreprise familiale au sein de l'entreprise. En France, nous pourrions retrouver ce type de culture chez les cheminots ou bien dans l'entreprise Michelin pour ne citer que deux entreprises nationales ou privées concernées par ces pratiques entrepreneuriales. Des commissions Américaine, bien sûr, mais aussi Anglaise, Ecossaise, Allemande, Suisse, Finlandaise et enfin Française sont constituées. Les résultats de cette enquête internationale ne font pas l'objet de publications particulièrement marquantes, sauf de la part de la France et de l'Angleterre. Le rapport français publié en 1936, nommé « *La correction des épreuves écrites dans les examens* », sera celui qui restera dans toutes les mémoires en regard de sa portée historique.

Le travail de la commission Française portait sur une expérience de multi-corrections des épreuves du baccalauréat. Cent copies furent tirées des archives, elles concernaient des épreuves suivantes : une dissertation philosophique, une composition française, une composition de langue anglaise, une version latine, une composition mathématique et une composition de physique. Nous pouvons retrouver cette étude dans beaucoup d'ouvrages ; néanmoins, la grande disparité de données et d'interprétations nous amène à rapporter la version la plus à même de retranscrire fidèlement la réalité des choses. Dans ce contexte,

³³ De Ketele JM. (1993) : « *L'évaluation conjugée en paradigmes* ». In: Revue française de pédagogie. Volume 103, 1993. P. 61

³⁴ Martin J. (2002) : « *Aux origines de la « science des examens » (1920-1940)* », p. 2

nous nous appuyerons sur la version de H. PIERON³⁵, le plus proche des auteurs, en sachant que le travail original de LAUGIER et WIENBERG ne nous a pas été accessible.

La méthode

Les recopies de ces copies furent transmises à six groupes de cinq correcteurs (examineurs). Par ailleurs, trois copies de composition française furent proposées à l'évaluation de soixante seize correcteurs. A ce sujet, nous pouvons remarquer le niveau d'investissement du corps enseignant et des expérimentateurs quant au nombre de personnes sollicitées pour ce travail. De plus, H. PIERON précise :

« (Les examinateurs) ... qui les notèrent et fournirent ensuite un rapport sur leur méthode propre de cotation, les qualités recherchées et leur hiérarchie, l'appel éventuel à des composantes distinctes, etc. »

L'interprétation des notions de notation et d'évaluation étaient donc libre, ce qui était courant à l'époque et très souvent encore maintenant. Seule l'attribution d'une note sur vingt points était exigée.

Résultats et discussions

L'analyse des données collectées a permis de dégager quatre éléments de réflexion selon LAUGIER et WEINBERG. Ceux-ci vont poser les bases de la docimologie et sont encore cités en références dès lors qu'il est fait allusion à cette science.

1- « L'écart des notes entre correcteurs »

Comme nous pouvons le constater sur l'illustration suivante, l'écart entre deux correcteurs est particulièrement important. Celui-ci va en moyenne de 3,29 à 3,36 en composition française et en philosophie, et de 1,88 à 2,05 en physique et en mathématique. L'écart maximum atteint des sommets en composition française et en philosophie avec des maxima à 13 et 12 points. Plus étonnant, nous pouvons constater des écarts maximums de 9 et 8 points en mathématique et physique, réputées être des sciences exactes, aux démonstrations précises et donc se prêtant peu à l'interprétation.

³⁵ Piéron H. (1963). « Examens et Docimologie » pp.19 à 24

	Ecart moyen	Ecart le plus fréquent	Ecart maximum
Composition française	3,29	6 et 7	13
Version latine	2,97	5	12
Anglais	2,24	4	9
Mathématiques.....	2,05	4	9
Philosophie	3,36	5 et 7	12
Physique	1,88	4	8

Figure 1 : Ecart des notes entre correcteurs. D'après H. PIERON³⁶

Il est à noter que selon H. PIERON, l'accord concernant les notes inférieures à dix sur vingt est plus fréquent que pour celles égales ou supérieures à la moyenne (30% vs 13%). Nous pouvons remarquer qu'aucune étude statistique ne vient corroborer une différence significative entre les deux groupes. L'impossibilité de récupérer les données initiales ne nous a pas permis de réaliser un test statistique complémentaire propre à confirmer ce résultat.

2- « Les échelles de notes »

Contrairement à ce qui est communément admis en statistique, l'échelle des notes ne prend pas l'allure d'une courbe normale³⁷, mais varie d'un correcteur à l'autre. Ces courbes varient de très étalées à très pointues signifiant des notes largement distribuées dans le champ des possibles ou bien centrées autour d'une moyenne propre au correcteur attestant de son penchant vers l'indulgence ou la sévérité. Certain de ceux-ci pouvaient également se centrer autour de la moyenne théorique et être qualifiés d'indécis. Centration et dispersion restent des notions statistiques qui ne réfèrent en aucune manière à la compétence du correcteur à effectuer la tâche qui lui est attribuée.

3- « Intercorrélations entre examinateurs »

L'intercorrélation est une notion particulière à appréhender. Pour résumer, nous pourrions dire qu'il y a intercorrélation si deux correcteurs n'ayant pas la même échelle de notation s'accordent sur le classement des différentes copies, à savoir de la meilleure à la plus mauvaise. Selon LAUGIER et WIENBERG, les divergences de corrélation *sont* « des divergences effectives qui peuvent être considérables ». Là encore, la démonstration s'appuie sur des comparaisons de moyennes et ne peut être représentative d'une différence effective en dehors d'un test statistique même si les valeurs calculées semblent en accord avec les conclusions des auteurs. Ces critiques statistiques ne sont, comme pour le

³⁶ Ibid. p 20

³⁷ Au sens d'une distribution selon la loi normale de Gauss et Laplace.

paragraphe précédent, que d'ordre méthodologique, l'avenir montrera la fiabilité des termes énoncés.

4- « La « vraie note » et la précision des correcteurs »

La grande variabilité des notations a amenée les auteurs à se poser la question de ce que pourrait être « la vraie note ». L'idée initiale a été de penser qu'il fallait, à l'instar des mesures physiques, augmenter le nombre de correcteurs (donc le nombre de « mesures »). Ce nombre devait être suffisant pour que la répartition des notes se distribue selon une courbe « en cloche » typique de la Loi normale de Gauss et Laplace. A cette condition, la moyenne calculée serait représentative d'un consensus d'accord du panel de correcteurs³⁸. L'outil statistique utilisé était la formule de Spearman-Brown dont l'usage premier est la recherche d'amélioration d'un instrument de mesure. Les résultats obtenus sont les suivants :

Composition française.....	78
Version latine.....	19
Anglais.....	28
Mathématiques.....	13
Dissertation philosophique.....	127
Physique	16

Figure 2: Nombre de correcteur pour une "vraie note" D'après H. PIERON

127 correcteurs pour une dissertation philosophique, 78 pour une composition française et encore 13 pour une épreuve de mathématique, pourtant considérée comme particulièrement rigoureuse, seraient nécessaires pour proposer une « note vraie »!

Le pavé est jeté dans la mare et ses vagues agitent encore aujourd'hui le monde de l'enseignement. D'aucun l'ignore, d'autre le réfute, ce résultat ne laisse personne indifférent, si ce n'est une proportion assez importante de correcteurs qui n'ont pas connaissance de ces travaux.

Issu de l'arsenal mathématique des physiciens, la formule de Spearman-Brown a été critiquée pour sa propension à attribuer une masse physique à un étudiant ou une copie. A ce sujet nous pouvons remarquer qu'il s'agit là d'une pratique courante dans le champ de la psychologie expérimentale et qu'en aucun cas les auteurs n'ont fait l'amalgame entre une copie ou un étudiant et un vulgaire kilogramme de métal ou de sucre.

³⁸ Sachant que dans ce cas, l'adjonction de notes supplémentaires ne fera plus varier la moyenne.

Parmi les critiques formulées à l'encontre de l'utilisation de la formule de Spearman-Brown, J-J. BONNIOL³⁹ en 1981⁴⁰ relayé par F. DIXIMIER⁴¹ en 2010, présente une critique de ces résultats statistiques et soulève une interprétation erronée de l'intervalle de confiance. La conséquence serait une augmentation sensible du nombre de correcteurs (762 pour une épreuve de philosophie !). Tout ceci ne change rien au problème mais lui donne un contraste encore plus caricatural, mettant l'accent sur l'impossibilité de réaliser un tel type de correction. Parallèlement, J-J. BONNIOL reprends l'idée qu'une « *copie n'est pas un objet physique ... au sens de la mesurer, mais de l'apprécier en référence à une échelle de valeurs* ». Le débat reste donc ouvert, mais la docimologie se présente encore malgré tout comme « un pied dans une porte entre-ouverte » sur la remise en question des pratiques des évaluateurs.

Pour terminer ce panorama de la « vraie note », LAUGIER et WEINBERG ont étudié les résultats de la seconde partie de l'enquête concernant la correction des trois copies de composition française par soixante seize correcteurs. Si la répartition des notes suivait une loi Gaussienne et permettait le calcul d'une moyenne, la dispersion de l'échantillon était énorme ce qui fait dire à H. PIERON :

« *Pour prédire la note d'un candidat, il vaut mieux connaître son examinateur que lui-même.* » et termine en citant P.BARRIER : « *Les correcteurs ... réagissent selon leurs goûts propres et leur conception personnelle de la culture secondaire* »

2.2.2. A ce propos, l'examineur se connaît-il lui-même ?

Pour répondre à cette question, nous pouvons rappeler une expérience de LAUGIER et WIENBERG concernant la possibilité d'une re-correction par le même examinateur quelques années plus tard. Ceux-ci ont demandé à un professeur universitaire de physiologie d'évaluer trente sept copies qu'il avait lui-même corrigées plus de trois ans auparavant. Seules sept copies reçurent la même note. Dans le cas des trente autres, des écarts de un à dix points sur vingt ont été relevés. Pensant qu'il pouvait s'agir d'un cas isolé, les auteurs ont répété l'expérience avec deux autres professeurs également reconnus compétents et expérimentés. Le constat initial s'est reproduit dans les deux cas. Ce travail

³⁹ Ancien professeur des universités. Fondateur et ancien directeur du département des sciences de l'éducation à l'Université de Provence, Aix-Marseille.

⁴⁰ Thèse « Déterminants et mécanismes des comportements d'évaluation d'épreuves scolaires » ; Université de Bordeaux II

⁴¹ Dixmier F. (2010). « *La vraie fausse note* » ; Pénombre – La lettre grise – www.olivier.hammam.free.fr/mports/penombre/lg7/lg07b.htm

est moins connu que l'enquête CARNEGIE, il fut d'ailleurs peu commenté, peut-être en raison du nombre restreint d'observations et centré sur une seule discipline. En 1968, la notion est reprise d'une façon encore plus provocatrice dans un travail paru dans « Science et Vie⁴² » et rapporté sur son site par Jacques NIMIER de l'IUFM de Reims :

« On demanda à une bachelière, Paulette, intelligente mais ignorant tout de la question traitée, de noter à son tour ces compositions de physiologie, après les avoir lues une fois pour se faire une idée du sujet. Ses notes eurent une corrélation de 0,51 avec celles attribuées par les professeurs compétents. La bachelière ne se trouvait pas plus en désaccord avec les spécialistes que ceux-ci entre eux. »

Ce travail, et d'autres sur le même thème, ont beaucoup ému la communauté des enseignants de l'époque. J. NIMIER parle même de choc affectif. La docimologie avait déjà froissé quelques susceptibilités de l'éducation nationale trente ans plus tôt, ce n'était pas cet article qui allait arranger les choses en ces temps troublés de « mai 68 ». Néanmoins, ceci permet de répondre en partie à la question titre de ce paragraphe. Si le correcteur reste conscient des valeurs qui le porte à poser un jugement sur une copie, l'évolution de ses convictions et du contexte de sa discipline ou de son environnement de travail l'amène à modifier ses représentations et donc son jugement professoral.

2.2.3. L'illusion de la double correction.

Au cours de cette étude nous avons pu constater que la pratique de la multiple correction portée à son extrême pouvait apporter une solution quant à l'attribution d'une note qu'on qualifierait aujourd'hui de « réaliste ». Moins réalistes sont en revanche les modalités d'une telle pratique. En effet comment réunir plusieurs dizaines, voire centaines, de correcteurs autour d'une même copie ? Ce constat pointe du doigt la pratique encore très courante de la double correction comme garante d'une note juste et bien équilibrée. A ce propos, l'établissement dans lequel se situe notre activité est soumis à une obligation réglementaire de la pratique de la double correction.

Au-delà de l'inaccessible atteinte de l'objectif visé, la double correction aura au moins pour avantage de donner au candidat, ignorant des arcanes de la docimologie, l'illusion d'avoir reçu une appréciation équitable du fruit de son travail.

⁴² N° 610 : Jacques Nimier, IUFM de Reims http://perso.wanadoo.fr/jacques.nimier/dossier_evaluation.htm

2.2.4. Conclusion

Le résultat de ces travaux est sans appel et posera dans les années qui vont suivre, et jusqu'à aujourd'hui encore, bon nombre d'interrogations; non pas quant-à la qualité, mais plutôt à la pertinence de la correction des copies par les examinateurs. A titre d'exemple, P. MERLE en 2010 intitule un chapitre d'un de ses articles⁴³ : « *la docimologie ou l'impossible exactitude évaluative* ».

La citation de H. PIERON en 1963, sur les bases de l'enquête de 1936, reste étonnamment moderne et d'actualité concernant l'état d'esprit des enseignants et formateurs en charge de la correction d'épreuves :

« ... si l'on décidait enfin de préparer à leur tâche délicate les examinateurs, et il serait souhaitable que pour chacune des matières envisagées, les spécialités appelées à les noter puissent y trouver matière à méditation, car les problèmes posés sont toujours actuels. »

A ce titre P. MERLE reprend cette idée en considérant le manque de prise en compte de ces données par les correcteurs, pointant de fait le manque de remise en question des enseignants. Peut-être pourrions-nous rajouter, leur manque de formation ? Il nous faut préciser que cet état est plus à mettre au crédit d'une méconnaissance de la docimologie que d'un refus des constats de celle-ci.

2.3.Phase deux : Du constat à ses causes, l'évaluation de la note et ses biais

Si la première partie de l'histoire de la docimologie a été celle du constat, la seconde s'attachera plus aux causes, et mettra en évidence les biais inhérents au correcteur, responsable du constat initial.

Il ne s'agit pas dans notre travail de faire un catalogue exhaustif des biais de l'évaluation, mais plutôt de mettre en avant les éléments essentiels à la compréhension du travail qui va suivre. Nous tenons à préciser que tous ces biais sont des notions qualifiées du nom « d'effets » par leurs différents auteurs.

⁴³ Merle P. (2010) : « *L'évaluation en philosophie : quels problèmes ?* » In « Côté-Philo » n° 14

2.3.1. Notion d'effet de tendance centrale

Basé une fois encore sur la Loi de GAUSS, l'effet de tendance centrale postule que plus le nombre de copies est important, plus un correcteur aura tendance à centrer son évaluation autour d'une valeur de note à peu près constante. Le corolaire de cet effet constaté chez les examinateurs, est de considérer grossièrement qu'il existe de façon constante un tiers d'élèves moyens, un tiers de bons et un tiers de médiocres. La réalité de la courbe de GAUSS est sensiblement différente comme le fait remarquer G. DE LANSHEERE⁴⁴ :

« Dans les sciences humaines, la courbe en cloche de Gauss joue un rôle considérable, parce qu'elle est l'image même de la répartition de bien des aptitudes et des qualités : les individus moyens abondent, mais les génies et les idiots, les géants et les nains sont rares. Comme les tests mesurent souvent des aptitudes, des traits de personnalité ou des performances de vastes populations, et servent à classer les individus en les comparant les uns aux autres, il est naturel que ces épreuves soient étalonnées selon la répartition gaussienne : en gros, 70 % de moyens, 13 % de bons, 13 % de médiocres, 2 % d'excellents, 2 % de très mauvais. »

Nous pouvons donc considérer qu'il est possible de répartir les examinateurs en trois catégories :

- Les « sévères » dont la tendance centrale sera inférieure à la moyenne théorique de l'épreuve,
- Les « laxistes », qui à l'inverse des précédents centreront leur notation au dessus de la moyenne théorique,
- Les « timides », qui de crainte de surévaluer ou sous-évaluer, auront une tendance centrale située au plus près de la moyenne théorique.

Ces dernières hypothèses seront confirmées par un enseignant hollandais du nom de POSTHUMUS en 1947 dans la loi qui porte son nom et formulée ainsi :

« Un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver, d'année en année, approximativement la même distribution (gaussienne) de notes. »

⁴⁴ De Landsheere G. (1971). « *Evaluation continue et examens. Précis de docimologie* »

Cette notion avait déjà été remarquée en 1927 par LAUGIER et WEINBERG avant leur participation à l'enquête CARNEGIE, mais cette fois-ci, la Loi Posthumus va plus loin, en parlant d'ajustement non seulement des appréciations, mais aussi des enseignements. Cette notion très importante de tendance centrale sera reprise dans les parties suivantes de notre travail.

2.3.2. Notion d'effet de stéréotypie

Ce biais de correction peut s'assimiler à une catégorisation d'un étudiant par l'examineur dès la première note posée. Il en résulte une notation spécifique à l'étudiant considéré, qui se verra attribuer des notes systématiquement centrée sur le niveau initial constaté par le correcteur. Cet effet bien connu des enseignants (et des élèves !) des cycles primaires et secondaires, tend à se restreindre lors des épreuves écrites anonymes des examens et concours. L'anonymat, spécifique de ces épreuves, est donc une bonne parade à ce biais, mais ne peut bien entendu être utilisable pour une épreuve orale.

2.3.3. Notion d'effet ROSENTHAL ou « effet PYGMALION »

Robert ROSENTHAL est un psychologue Américain d'origine allemande. Dans la droite ligne de la psychologie expérimentale, R. ROSENTHAL a réalisé l'expérience suivante en tenant compte des travaux de Burrhus Frédéric SKINNER, fondateur du mouvement behavioriste radical initié par Ivan PAVLOV. Par ailleurs, à la suite de ses travaux, B-F SKINNER a établi le concept de conditionnement opérant.

Dans un premier temps l'expérience a porté sur deux groupes de rats présentés pour le premier comme une sélection d'animaux choisis pour leur intelligence exceptionnelle, et le second comme un lot tout venant de rongeurs sans aptitudes cognitives particulières. Deux groupes de six étudiants prirent en charge ces deux lots d'animaux et les soumirent au test de la boîte de SKINNER (voir figure page suivante) utilisée initialement dans les expériences de conditionnement opérant de son auteur. L'objectif de cette étude était de déterminer les capacités d'apprentissage des animaux dans ces conditions. Les deux groupes de rats avaient été constitués de façon aléatoire sans aucune notion de capacités cognitives particulières, les expérimentateurs n'en étaient bien entendu pas prévenus.

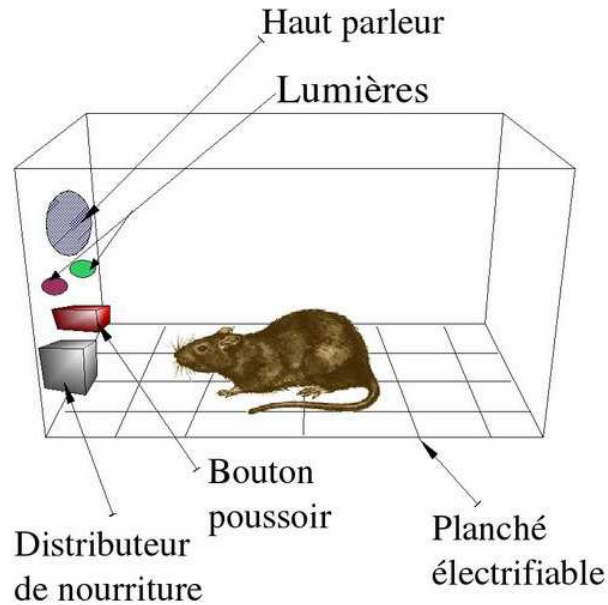


Figure 3: exemple de boîte de SKINNER⁴⁵

Le groupe des « sympathiques rongeurs » étiquetés intelligents a vu ses résultats dépasser très largement celui des « rats ordinaires ». Les raisons de cet écart pourraient être débattues longuement, mais ce n'est pas le sujet de notre travail. Nous nous contenterons de nous borner aux faits démontrés.

Le second temps de l'étude a consisté en une vérification de l'hypothèse selon laquelle l'attention que l'on porte à quelqu'un va conditionner la réussite de celui-ci. La transposition de l'expérience s'est donc faite des rats vers des enfants en cours de scolarité (même si les enfants n'ont pas été soumis au test de la boîte de SKINNER.). Ce travail a été mené en collaboration de Leonore JACOBSON dans une école de San Francisco.

Le scénario de l'étude a été le même. Deux groupes d'enfants choisis au hasard, mais néanmoins testés auparavant sur le plan intellectuel, ont été présentés à leurs professeurs. L'un était réputé constitué d'enfants brillants promis à un bel avenir intellectuel et professionnel, et l'autre d'élèves très ordinaires, sans aptitudes particulières⁴⁶. La boîte de SKINNER était pour la circonstance remplacée par l'école qu'ils fréquentaient. L'évaluation de leurs capacités cognitives et intellectuelles fut réalisée auprès de ces élèves au bout d'une année sous la forme d'un test non verbal d'intelligence générale.

⁴⁵ http://fr.wikipedia.org/wiki/Bo%C3%AEte_de_Skinner

⁴⁶ Pour renforcer l'effet d'annonce, Rosenthal et Jacobson ont préféré fournir un courrier « *attribué par erreur* » aux enseignants. Celui-ci faisait état de pseudos résultats obtenus par les élèves aux tests de QI favorisant certains plus que d'autres de façon aléatoire.

Les résultats furent éloquentes : une augmentation de cinq à vingt cinq points d'indice de quotient intellectuel fut constatée dans le groupe présenté comme brillant. La différence avec le groupe contrôle fut jugée nettement significative.

Le récit et la présentation de ce travail a fait l'objet d'un ouvrage cosigné R. ROSENTHAL et E. JACOBSON. Son titre, « Pygmalion à l'école »⁴⁷ à transformé le terme d'effet Rosenthal en effet Pygmalion⁴⁸. Nous pouvons remarquer à ce propos qu'il s'agissait du concept de « *prophétie réalisatrice* »⁴⁹ énoncé initialement par R. ROSENTHAL lors de sa première étude.

Généralement classé dans la catégorie des biais, l'effet Rosenthal ne serait-il pas plutôt un bien ?

A titre personnel, nous pensons qu'il devrait plutôt faire l'objet d'un enseignement spécifique auprès de tous les enseignants et formateurs. En effet, comment ne pas prendre en compte nos représentations des candidats et futurs étudiants orientées vers une croyance en leurs capacités ? Si la réussite aux formations dépend de la confiance que nous accordons à leur potentiel, il ne nous paraît pas opportun de s'en affranchir. Cette notion tranche avec celle de renforcement positif, couramment enseignée dans les écoles de management. En effet cette dernière tend plus vers l'exploitation d'un égo flatté que vers une recherche de l'épanouissement personnel induit par la confiance.

2.3.4. Notion d'effet de contraste (ou effet de comparaison ou effet d'ordre de correction)

Ce biais est particulièrement bien connu des élèves et des étudiants. Qui n'a jamais dit : « en étant corrigé (ou examiné à l'oral) juste après cette personne (la « première de la classe ») il est bien normal que j'obtienne une aussi mauvaise note ». Il est tout à fait compréhensible qu'un correcteur ou un jury soit sensible à cet effet, c'est une perte de repères des valeurs à attribuer aux étudiants. La conséquence de ceci risque d'être une sous estimation majeure d'une prestation moyenne ou, a contrario, une surestimation extrême si le candidat précédent était particulièrement mauvais. Nous pourrions comparer cet effet à la période d'adaptation nécessaire à la vision humaine lorsque nous passons de la lumière

⁴⁷ Rosenthal R., Jacobson L. (1978) : « *Pygmalion à l'école* ». Paris : Ed. Casterman

⁴⁸ Le nom de Pygmalion fait référence à la mythologie grecque. Pygmalion était un artiste qui tomba amoureux d'une de ses sculptures féminines. Nous pouvons y voir le symbole d'un aveuglement qui fait perdre la notion de réalité. La légende de Pygmalion et Galatée est aussi l'incarnation de la puissance de l'imaginaire dont l'évocation parvient à tromper son créateur jusqu'à la rendre réelle.

⁴⁹ « Self-fulfilling prophecy » dans le texte original

vive du soleil à la pénombre d'une maison. Dans un premier temps notre idée sera qu'il fait particulièrement sombre puis, progressivement, tout paraîtra normal. Cette analogie ne s'arrête pas là, elle met l'accent sur la rémanence de l'effet qui demande un laps de temps suffisant pour disparaître : ce n'est donc pas un, mais plusieurs candidats ou étudiants qui pourront être affectés par ce biais de contraste. Dans ce cas précis, l'effet de contraste est lié à l'effet d'ordre.

Il existe une variante à cet effet de comparaison appelé « effet d'ancrage ». Dans ce cas, une copie ou une prestation sera prise en référence et toutes les autres notes seront attribuées en fonction de celle-ci. A notre sens, il ne s'agit pas réellement d'un biais, mais plutôt d'une méthode de travail permettant une certaine harmonisation de la notation d'ensemble. Le revers de cette méthode est le risque d'accentuer l'effet de tendance centrale si le repère initial représente une prestation très proche de la moyenne.

Une petite nuance peut être apportée à l'effet d'ordre de correction. Celui-ci serait à rapprocher d'un « effet de lassitude » considérant que les premières copies d'un groupe bénéficieraient de plus d'indulgence de la part du correcteur que les dernières⁵⁰. L'idée est recevable mais tout cela dépend de l'état d'esprit dans lequel est entrepris le travail. En effet, si l'on considère un examinateur d'humeur maussade commençant sa pile de correction par une copie particulièrement brillante, la tendance peut très rapidement s'inverser. L'effet de contraste sera alors au bénéfice de l'ensemble du groupe, si tant est que la tendance « maussade » du dit correcteur ne soit pas une constante psychologique majeure. Dans ce cas, l'effet de tendance centrale à mode nettement inférieur à la moyenne sera respecté.

Nous ferons remarquer qu'il existe un moyen simple de s'affranchir, ou plutôt d'amoindrir, le risque d'effet de contraste lors de la correction de copies comprenant plusieurs questions. Cette méthode consiste en une correction que nous qualifierons de séquentielle, c'est-à-dire d'évaluer la même question pour toutes les copies, puis de modifier l'ordre des copies pour la correction de la question suivante, ainsi de suite jusqu'à la dernière question. Si ce principe ne supprimera pas totalement le contraste entre les copies, il le diminuera de façon très significative.

⁵⁰ JJ. BONNIOL 1978 in « Psychologie de l'évaluation scolaire »

2.3.5. Notion d'effet de halo

L'effet de halo n'est pas spécifique de la docimologie. En 1920, Edward THORNDIKE⁵¹ a démontré qu'une caractéristique particulière ou une première impression favorable d'un individu va contaminer la perception cognitive que l'interlocuteur aura par la suite de cette personne. Cette vision parcellaire très positive sera généralisée à toutes les autres qualités attendues. C'est ainsi qu'une personne au physique agréable et à l'élocution aisée lors d'un entretien, se verra attribuer a priori des qualités d'intelligence, de compétences ou de rigueur sans avoir à en faire la démonstration. A l'inverse, quelqu'un ressenti de prime abord comme désagréable aura beaucoup de difficulté à faire valoir ses qualités à niveau effectif. G. DE LANDSHEERE⁵² l'exprime de la façon suivante :

« L'effet de halo présente un caractère affectif accusé. Souvent, on surestime les réponses d'un élève de belle allure, au regard franc, à la diction agréable ... Soit pour des raisons de lisibilité, soit pour des raisons nettement affectives, l'écriture peut aussi influencer le correcteur. »

Cela va introduire des biais de jugement important lors de l'évaluation des qualités recherchées par un jury ou un examinateur. Dans le cas d'un jury, l'effet peut être amoindri par le nombre de membres le constituant.

L'effet de halo est également très connu dans le monde du commerce, il est responsable entre autre de l'image que nous nous faisons d'une marque en attribuant à ses produits des qualités supposées qui n'ont jamais été montrées⁵³.

2.3.6. L'arrivée des sociologues et la mise en évidence de la notion de biais sociaux.⁵⁴

La docimologie reste le domaine privilégié des psychologues. Cette constatation n'est pas nouvelle et fait référence à la méthodologie de travail la plus couramment utilisée, à savoir la méthode expérimentale. Déjà en 1932, H. PIERON évoque dans le *Bulletin de l'Institut National de l'Orientation Professionnelle*⁵⁵, dont il est le principal rédacteur :

⁵¹ (1874-1949) Psychologue de l'éducation américain.

⁵² De Landsheere G. (1971). « *Evaluation continue et examens. Précis de docimologie* » p. 49

⁵³ Dans les années soixante, une étude avait montrée que les acheteurs considéraient que la marque allemande GÜNDIG fournissait les meilleurs électrophones. A cette époque là, cette marque n'avait jamais fabriqué d'électrophone !

⁵⁴ 1996 « *L'évaluation des élèves, enquête sur le jugement professoral* : » in : in : Revue française de pédagogie, Année 1998, Volume 122, Numéro 1 p p 13 – 15 et 21

⁵⁵ BINOP novembre 1932 p 238

«... certaines résistances à l'introduction du point de vue psychotechnique, résistances que l'on s'est habitué à rencontrer chez quelques sociologues de tendances littéraires et qui semblent redouter l'introduction de l'esprit expérimental »

Jusqu'ici absents dans les débats docimologiques, les sociologues apportent leur pierre à l'édifice de la docimologie, en particulier P. MERLE⁵⁶. Ils mettent en lumière des biais sociaux de l'évaluation concernant les catégories sociales, le genre, le parcours scolaire préalable et l'âge.

Tout d'abord, un élève redoublant ou ayant redoublé, identifiable par son âge plus élevé sera noté de façon plus sévère qu'un élève faisant état d'un parcours scolaire sans embuches.

Le sexe est également pris en compte par les chercheurs qui montrent que l'évaluation notée des filles est en moyenne plus élevée que celle des garçons. L'observation reste pertinente même lorsque les résultats aux tests anonymes montrent des niveaux d'aptitudes identiques.

De même « les enfants de cadres supérieurs sont mieux notés que les enfants des autres milieux » selon P. MERLE qui cite les travaux de Bella et Mingat en 1993. Nous retrouvons également cette tendance dans les appréciations portées en regard d'une note : à note égale, un élève issu d'une famille modeste recevra à note identique une appréciation plus dévalorisante⁵⁷.

Nous pourrions constater que le recours à l'anonymat est encore une fois l'occasion d'éviter ces biais sociaux lors des épreuves écrites. Il est bien entendu qu'il ne s'agit pas de porter un jugement sur l'intégrité intellectuelle des examinateurs, mais de faire le constat de la présence de ces différents biais.

L'objectif de ce chapitre n'est pas de faire un panorama exhaustif de tous les biais de l'évaluation, mais de préciser ceux, qui, par leurs importances, pourront apporter un éclairage à la connaissance de la docimologie et à son apport aux sciences de l'éducation. Certaines des notions précédemment définies seront utilisées dans la suite de notre travail.

Parallèlement à l'émergence de la prise en compte de ces nouveaux biais de l'évaluation et de la notation, la docimologie retourne depuis le début du nouveau millénaire à la

⁵⁶ Op. cit.

⁵⁷ Daniel ZIMMERMAN en 1982 cité par P. MERLE

recherche d'éléments complémentaires propres à confirmer et à consolider ses fondamentaux.

2.4.Phase trois : un retour aux fondamentaux et l'amorce de propositions d'améliorations

2.4.1. Le commentaire à la note : une notion à améliorer

De la fin des années soixante à la fin des années quatre-vingt, les chercheurs en docimologie s'attacheront surtout à la forme de la correction des copies de l'enseignement primaire et secondaire. Les commentaires associés à une correction de copie seront surtout étudiés, en particulier, l'incidence de ceux-ci sur les apprentissages. Il ressort de ces multiples travaux⁵⁸ qu'un apport méthodologique, sous forme de guides de corrections, serait à même de diminuer les commentaires empiriques des examinateurs. Ces divers commentaires étant souvent de nature à perturber la compréhension du sujet par l'élève au profit d'un ressenti négatif associé à une brimade.

2.4.2. Bruno SUCHAUT :Une synthèse du retour aux fondamentaux

Plus près de nous, il faut citer le remarquable travail de B. SUCHAUT⁵⁹ en 2008. Celui-ci présente trois parties distinctes.

Dans un premier temps, l'auteur fait la synthèse des travaux docimologiques entrepris depuis le début des années soixante-dix jusqu'à nos jours. Il ressort de cette revue de la littérature qu'hormis la mise en évidence des biais sociaux présentés dans le chapitre précédent, peu de changement ont pu être observés depuis les travaux issus de l'enquête CARNegie.

La deuxième partie du travail consiste en une expérimentation sur les copies baccalauréat en multi correction. La méthode n'est pas sans rappeler, dans l'esprit,, les travaux de LAUGIER et WEINBERG soixante-douze ans plus tôt. La différence méthodologique se situe dans l'orientation statistique initiale du choix des copies et de l'origine géographique⁶⁰ des correcteurs. Cette particularité offre la possibilité de comparer entre

⁵⁸Rapportés par C.CAPELLE (2010) : « *Pratiques de correction sur copies d'examen et nouveaux usages instrumentés* » pp.3 -- 10

⁵⁹ Suchaud B. (2008) : « *La loterie des notes au bac. Un réexamen arbitraire de la notation des élèves* » Les Documents de Travail de L'IREDU (Institut de Recherche sur l'EDUcation)

⁶⁰ Issus respectivement des académies de Besançon et de Dijon.

eux deux groupes distincts d'examineurs et d'établir si les différences observées sont significatives ou non. L'utilisation d'un arsenal plus complet d'outils statistiques performants⁶¹ a permis de confirmer les écarts intra-groupes constatés en 1936 et de constater que ces écarts dans les deux groupes sont comparables. La recherche des causes possibles de ces variations par l'analyse qualitative des commentaires associés aux copies fait apparaître une grande diversité difficile à analyser. Néanmoins, il faut remarquer d'un point de vue quantitatif :

« ... une corrélation négative et significative entre le nombre de mots composant les commentaires et les notes. C'est comme si le correcteur se sentait d'autant plus contraint, même dans un cadre expérimental, de justifier sa pratique quand son jugement sur la production de l'élève est négatif.⁶² »

La troisième partie de l'étude fait état des pistes de réflexions susceptibles d'améliorer un constat particulièrement défavorable au système actuel d'évaluation notée des performances des étudiants soumis à des épreuves pouvant être qualifiées de déterminantes pour leurs avenir :

Tout d'abord, une fois encore, fournir des outils tels que des grilles de corrections propres « à renforcer la cohérence des pratiques » et « garantir des aléas moins importants » en termes de notations semble être une solution souhaitable. Ensuite, privilégier l'usage de Questions à Choix Multiples (QCM) pour réduire les marges d'interprétation des réponses. Nous remarquerons à cet égard que l'auteur parle de « marges d'interprétation nulles » ce qui nous paraît un peu excessif au regard des erreurs possibles des examinateurs, au long terme, sur ce type de corrections particulièrement répétitives. La seconde étape de réflexion se situe au niveau de la remise en question de la nécessité d'existence du baccalauréat, et par extension des examens, en termes de coût, de valeur et de sélectivité. La notation est également remise en question dans cette étude, en remarquant le peu de crédibilité constaté depuis près de quatre vingt ans, ainsi que le peu d'intérêt qui lui est accordée dans bon nombre de pays, en particulier en Europe du nord.

Ces pistes, de l'ordre de la stratégie et de l'ingénierie, soulignent l'ampleur du travail à accomplir en matière d'examen et d'évolution des mentalités des examinateurs ou correcteurs, ce qui fait dire à l'auteur : « faire le deuil de la notation, renvoie aussi à changer plus globalement la finalité de l'acte d'enseignement. »

⁶¹ Analyse de la variance par test ANOVA et test t de Student

⁶² Suchaud B. (2008) : « La loterie des notes au bac. Un réexamen arbitraire de la notation des élèves » p.10 et 11

B. SUCHAUT n'est pas le seul à avoir revisité récemment les fondamentaux de la docimologie. Sur un autre plan, celui de l'étude des biais cognitifs de l'examineur, André ANTIBI nous propose une critique moderne de l'effet de tendance centrale.

2.4.3. La « constante macabre »

André ANTIBI représente un cas particulier des auteurs s'intéressant à la docimologie. Son parcours de Professeur de mathématique à l'Ecole Supérieure d'Aéronautique de Toulouse ne le prédestinait pas à l'intérêt qu'il porte actuellement à la docimologie en tant que Docteur en Sciences de l'Education. Suite à la publication de son livre « *la constante macabre* » en 2003⁶³, A. ANTIBI se trouve en 2005 devant la commission culturelle du sénat. La commission lui demande de bien vouloir définir cette fameuse « constante macabre » dont le fondement s'inscrit une fois de plus dans l'analyse contextuelle de la loi de GAUSS. Voici sa réponse inscrite au procès verbal d'audition⁶⁴ :

« Tout d'abord, M. André Antibi a donné une définition de cette « Constante macabre », correspondant au pourcentage constant d'élèves qui doivent être en situation d'échec pour que notre système d'évaluation soit crédible. »

L'explication est claire : si un correcteur ne permet pas un minimum d'échec à un examen ou une épreuve d'évaluation, sa crédibilité s'en trouvera affectée face à la profession et à sa hiérarchie. Toujours d'après A. ANTIBI, chaque enseignant est soumis à cette constante. Pratiquement, quelle que soit sa spécialité (excepté les arts plastiques et l'éducation physique) un enseignant l'applique de façon « *sournoise*⁶⁵. » Selon lui il s'agit d'une spécificité Française dont tous les enseignants sont conscients et à laquelle ils désirent mettre fin. La réalité montre un autre visage : celle, véritable, de la constante macabre.

Il s'agit là d'un exemple concret de l'effet de tendance centrale, caricatural mais néanmoins réel, lié à l'image que peut avoir un examinateur vis-à-vis de ses pairs et de son administration de tutelle.

La démotivation engendrée par une telle pratique doit, selon l'auteur, être combattue. En cela il rejoint P. MERLE qui préconise :

⁶³ Antibi A. (2003) : « *La Constante macabre* », Toulouse : Ed. Nathan Col. Math'Adore,

⁶⁴ Rapport de la commission culturelle du sénat de M. André ANTIBI le 05 décembre 2005

⁶⁵ Expression empruntée à A. ANTIBI

« ... l'aléa considérable de la notation ... devrait amener les professeurs à considérer avec plus de circonspection leurs notations.⁶⁶ »

2.4.4. En résumé de cette troisième phase

Les travaux réalisés par Bruno. SUCHAUT et Jean. ANTIBI sont très représentatifs de la docimologie moderne. Par les méthodologies utilisées, ils confirment l'orientation méthodologique expérimentale de la discipline par comparaison de groupes constitués. L'utilisation d'outils statistiques sur des valeurs quantitatives situe ce type de recherche dans le champ de l'édumétrie, issue de la psychologie expérimentale des origines et de la psychométrie, encore confidentielle en France, comparativement aux autres pays très actifs dans le domaine docimologique, que sont la Belgique, la Suisse et le Canada.

2.5. En conclusion : Qu'en est-il aujourd'hui, de la docimologie ?

Née au début du vingtième siècle, la docimologie a connu un essor important jusqu'à la fin des années trente. S'en est suivie une période de « sommeil » de vingt ans, entre la fin des années trente et le milieu des années soixante, marquée par une maigre production de travaux essentiellement en France. La seconde guerre mondiale, puis la reconstruction du pays peuvent expliquer cette constatation. En effet, la docimologie n'était certainement pas la priorité de cette époque troublée. Les années qui s'en suivirent ont permis de développer de nouveaux concepts dont les prolongements réflexifs sont encore d'actualité.

La docimologie est maintenant une science bien définie par les notions et les concepts qu'elle a mis en évidence et développés. Néanmoins, celle-ci cherche encore à élargir son audience auprès des acteurs de l'enseignement. L'explication que nous pourrions fournir se situe dans la particularité qu'a la docimologie à obliger les examinateurs, les correcteurs et les enseignants ou formateurs à remettre en questions leurs pratiques. En effet, à l'heure de la pratique réflexive, les enseignants et formateurs n'intègrent pas toujours ce concept de docimologie. Albert ENSTEIN fournit une hypothèse d'explication de cet état de fait lors d'une de ces citations connues : *« Il est plus difficile de désagréger un préjugé qu'un atome »*.

⁶⁶ Op cit p. 22

Pierre MERLE résume en ces termes « l'évolution » des correcteurs : « *la docimologie ou l'impossible exactitude évaluative* ». Cette réflexion met en évidence le sentiment d'impuissance des évaluateurs face à cette tâche quotidienne qui est la leur. L'émergence explicite d'autres questions dont on avait peu tenu compte durant les vingt années précédentes, se retrouvent sous la plume de Bruno SUCHAUT « *les notes du bac : une loterie* ». Encore une fois le sentiment d'impuissance, voire de fatalité, est mis en avant et les embryons de solutions donnent la mesure de la difficulté à changer des habitudes vieilles de plus de deux cents ans.

Au titre des propositions d'améliorations, nous remarquerons celle d'une prise en compte de la correction instrumentée par l'utilisation de l'outil informatique⁶⁷. A ce sujet, il est à noter que de telles propositions avaient été avancées au début des années soixante dix mais n'avaient pas été suivies d'effet faute de matériels et de logiciels suffisamment performants.

La notation doit-elle être supprimée⁶⁸ ? Ce problème épineux en France a été fréquemment débattu depuis 1969 jusqu'à la fin des années soixante dix et alimente encore aujourd'hui l'actualité de l'enseignement primaire. Des tentatives expérimentales ont été mises en place, mais aucune d'entre elles n'a apporté de solution. L'explication la plus souvent mise en avant tient dans les habitudes françaises centrées sur la sélection. Partant de ce principe particulièrement bien ancré dans notre pays, il faut donc bien continuer à noter les élèves ou les candidats afin de permettre un classement sensé sélectionner ceux qui sont aptes à entrer dans un établissement et à en suivre la formation !

Nous pouvons, dans ce contexte centré sur l'évaluation de l'étudiant poser la question d'une solution alternative :

Ne serait-ce pas plutôt dans le choix et la formation des enseignants, des formateurs et des examinateurs de tous poils, que dans la sélection des étudiants, que se trouverait la solution ?

La mise en évidence des biais de correction pose en effet ce type de question. Nous pensons en particulier à l'effet Rosenthal qui, s'il pouvait être maîtrisé, apporterait une réponse positive à cette interrogation dans le sens où celui-ci tire vers le haut la progression des étudiants. L'objectif de la réussite des étudiants dans leur futur professionnel n'est-il pas la raison même d'un enseignant ou d'un formateur alors qu'il a

⁶⁷C.CAPELLE (2010) : « *Pratiques de correction sur copies d'examen et nouveaux usages instrumentés* » p.11.

⁶⁸ De Landsheere V. (1971). « *Faire réussir, faire échouer : la compétence minimale et son évaluation* »

lui-même participé à leur sélection ? Il s'agit là encore d'un problème d'évolution des mentalités issu des représentations négatives du correcteur vis-à-vis du candidat ou de l'étudiant quant à ses capacités à réussir.

Sans avoir la prétention d'apporter des solutions définitives aux problèmes soulevés au cours de ce chapitre, souvent récurrents depuis une centaine d'années, nous nous proposons dans le chapitre suivant d'affiner la question de départ de ce travail afin d'en dégager une problématique susceptible d'être étudiée

Partie 2

II. Problématisation de la question de départ : étude détaillée de l'examen et compléments conceptuels.

1. LE CONSTAT DE DEPART : UNE QUESTION D'EFFICACITE DE LA SELECTION

Avant de porter plus avant notre réflexion, il nous semble important de rappeler la question initiale évoquée dans la première partie de notre travail. Il s'agit d'un questionnement concernant la sélection des candidats à suivre la formation spécialisée d'Infirmier Anesthésiste. La formulation était la suivante :

"Comment peut-on améliorer la sélection des candidats à suivre une formation lors d'un examen classant ou d'un concours ?"

Dans ce genre de questionnement, la première interrogation consistera en un bilan susceptible de confirmer ou d'infirmer la légitimité du questionnement. La question posée sera donc : Pouvons-nous affirmer que notre procédure d'évaluation nous amène à sélectionner de façon fiable des candidats aptes à suivre la formation ? Après un retour aux constats historiques, nous étudierons les résultats factuels servant de base à ce travail extraits de l'expérience des cinq dernières années de l'établissement concerné.

1.1. Un retour aux constats docimologiques historiques

Un retour historique s'impose pour poser les bases de notre réflexion, sachant que depuis cette époque peu d'éléments nouveaux à ce sujet sont à signaler. Dans son ouvrage de référence de 1963⁶⁹, H. PIERON fait état des constatations de l'étude CARNEGIE menée

⁶⁹ « La docimologie »

en France en 1936. La question posée interrogeait l'incidence des corrections sur l'avenir des candidats présentés à l'examen du baccalauréat ; a savoir admis ou reçus. Le tableau suivant présente les résultats de cette étude :

Disciplines	Candidats Admis par les six correcteurs	Candidats Refusés par les six correcteurs	Candidats Admis ou refusés
Mathématique	20%	44%	36%
Anglais	16%	37%	47%
Version latine	10%	40%	50%
Physique	13%	37%	50%
Composition Française	9%	21%	70%
Philosophie	10%	9%	81%

Figure 4 : candidats reçus ou refusés (selon H. PIERON 1963)

Très nettement, et ceci quelle que soit la matière envisagée, le pourcentage de désaccord (exprimé dans la colonne « Candidats admis ou refusés ») marque clairement la possibilité d'erreur relative, quant à l'évaluation d'admission ou d'ajournement d'un étudiant. Nous remarquerons que ces divergences sont d'autant plus marquées que la discipline se prête à une subjectivité de jugement, celles-ci vont de 36% en mathématique à 81% en philosophie.

Partant de ce constat historique, qui permet de se poser légitimement la question de la pertinence de notre jugement porté sur amissibilité, examinons ce qu'il en est de la fiabilité du type d'examen classant qui nous intéresse pour ce travail.

2. « LES PREMIERS SERONT LES DERNIERS⁷⁰ » ?

En dehors de toute considération biblique, l'objectif de cette première étude est de déterminer si l'ordre de classement au concours d'entrée correspond aux attentes de réussites à la fin du cursus de formation. Le choix d'un indicateur pertinent nous a conduit

⁷⁰ Evangile selon Saint Matthieu : « Ainsi les premiers seront les derniers et les premiers seront les derniers »

à nous intéresser à l'existence d'une possible corrélation entre le rang d'admission au concours d'entrée et le rang constaté à la fin de la période de formation, basé la sommes des notes attribuées pour l'obtention du Diplôme d'Etat. Cette dernière note ne pose aucune difficulté dans le sens où les étudiants présentés à l'épreuve finale peuvent aisément être classés par ordre de mérite (de façon officieuse bien entendu). Les ajournements sont extrêmement rares, trois en dix ans, dont un seul est comptabilisé dans les données collectées sur les cinq dernières années. Son admission a été prononcée secondairement lors d'une deuxième cession pour un défaut de validation de stage en fin d'année. Cette précision ne modifie en rien la validité de l'indicateur car ce type de non validation n'est pas assujetti d'une note. Le rang à l'admission pose plus de difficultés méthodologiques et nécessite d'être précisé.

2.1.La notion de rang relatif à l'admission

Il faut en effet considérer qu'un candidat peut être reçu sur la liste principale (quinze au maximum) ou sur la liste complémentaire. Dans le second cas, l'admission effective n'est prononcée que lorsqu'il est possible au candidat d'accéder à la liste principale à la suite de la défection d'un candidat de la liste principale pour différents motifs. Son rang au concours ne sera donc pas compris entre un et quinze comme le voudrait la logique⁷¹. De plus, un étudiant admis sur la liste principale peut demander auprès de la direction de l'établissement un report d'un an de son admission pour diverses raisons, en particulier défaut de financement, grossesse etc.

La difficulté à déterminer un rang d'admission pour la présente étude tient au fait qu'un candidat peut, par exemple entrer en formation avec un an de retard, dans le cas d'une demande de report, en étant classé premier au concours, et un autre, entrer la même année en étant classé vingt cinquième à la suite des épreuves par le jeu des désistements.

La méthode employée pour palier à cet inconvénient de classement consiste en un réajustement du rang réellement obtenu par chaque étudiant au concours (de un à vingt ou trente selon les années) ramené à un rang relatif de un à quinze étudiants effectivement présents lors de la rentrée. Cette méthode est à nos yeux suffisamment pertinente car elle conserve la mémoire de la valeur de l'évaluation de l'étudiant lors du concours. Elle permet de plus de pouvoir la comparer à celle des autres personnes de la promotion lors de l'examen final. Le corolaire d'un tel type de réajustement est d'admettre l'éventualité d'ex-

⁷¹Rappel : L'agrément de l'établissement fait état d'un maximum de quinze étudiants par année de formation.

æquo, ce qui est d'ailleurs le cas, sans aucune incidence méthodologique sur la suite de l'étude. Nous pouvons remarquer que, d'une part la possibilité d'obtenir des candidats ex-æquo diminue lorsque l'on s'éloigne des premières places, et que d'autre part, la probabilité que deux ex-æquo puissent simultanément obtenir des résultats diamétralement opposés à leur rang d'origine est extrêmement faible.

Nous ferons également remarquer qu'il n'y a pas le même nombre de premiers, de seconds ou de troisièmes etc. du simple fait que la possibilité de produire des ex-æquo exclue certains classements. Par exemple, s'il existe un premier et deux seconds, le prochain classé sera forcément compté comme quatrième. S'il n'en est pas ainsi, la possibilité de deux premiers et de deux seconds placerait le dernier à la dix-septième place alors qu'ils ne sont que quinze au maximum, ce qui constituerait une aberration.

2.2.L'échantillon étudié, le type d'étude et les outils utilisés

Le premier travail que nous présentons consiste à vérifier l'hypothèse selon laquelle le classement au concours d'admission, reflet de l'évaluation des aptitudes d'un candidat à suivre la formation, est respecté à la fin de celle-ci. Pour cela, nous avons réalisé une étude destinée à rechercher s'il existe une corrélation entre le rang relatif au concours d'entrée et le rang effectif du même individu à l'issue des deux années de formation. Cette étude est de type rétrospectif sur les données collectées des sept dernières années d'exercices dans l'établissement. Les résultats aux concours sont ceux des années 2005 à 2009. Pour les résultats aux Diplômes d'Etat, ils correspondent aux exercices 2007 à 2011 en raison du décalage de deux ans correspondant à la durée des études d'Infirmier Anesthésiste. Chaque étudiant est donc représenté par une paire de données représentatives de son rang à l'admission et de celui à la fin de sa formation⁷².

Les outils d'analyse informatiques utilisés sont le tableur Microsoft EXCEL® pour le recueil, le tri des données et les représentations graphiques courantes ainsi que le logiciel PRISM® pour le traitement de l'analyse statistique et la réalisation de certaines représentations graphiques spécifiques.

Nous allons présenter dans le chapitre suivant les résultats obtenus par cette étude. La première présentation sera de type graphique, descriptive, la seconde sera statistique. Nous conclurons par une discussion de ces résultats.

⁷² Voir annexe 1

2.3.Présentation descriptive graphique préliminaire

L'échantillon observé est constitué de soixante huit personnes représentant les cinq dernières promotions d'Infirmiers Anesthésistes Diplômés d'Etat sortis de l'établissement.

Nous pouvons remarquer que cela représente une moyenne de 13,6 étudiants par promotion soit un taux de remplissage de 91%.

Le graphique suivant présente en abscisse le rang relatif de l'étudiant au concours d'entrée et en ordonnée le rang à l'issue de la formation. Chaque point correspond à un étudiant et à son évolution sur deux ans en terme de classement. Pour cela nous avons présenté son rang relatif au concours d'entrée en abscisse (Rang concours relatif) et son rang de classement à l'issue de la formation, après les épreuves du Diplôme d'Etat, sur l'axe des ordonnées (Rang DE).

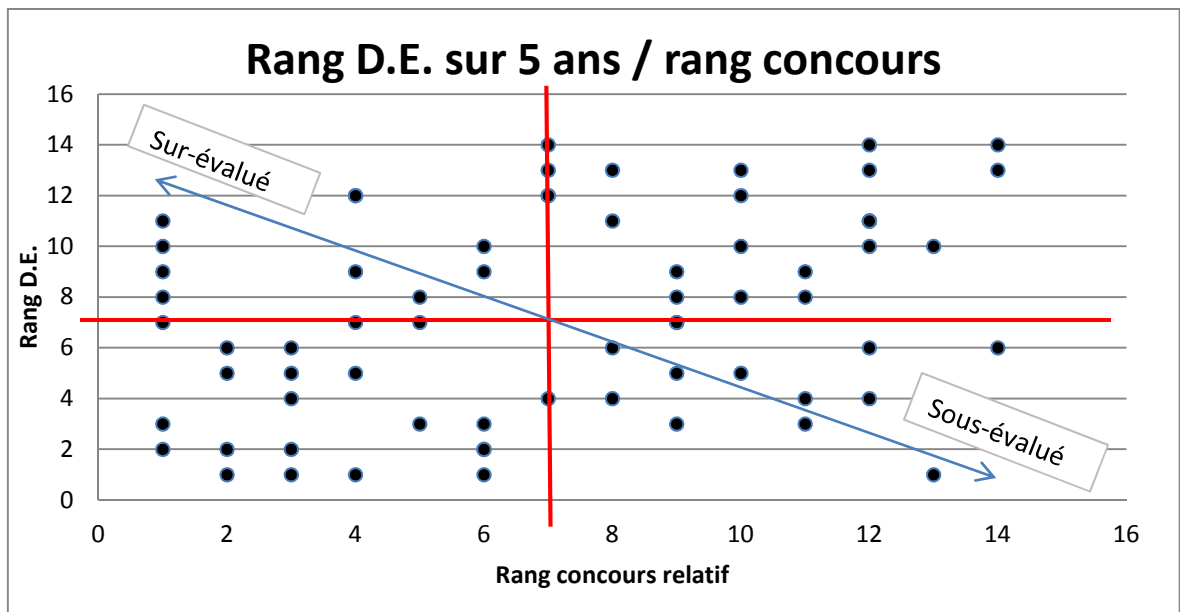


Figure 5: Rang au Diplôme d'Etat / Rang relatif au concours (sur 5 ans)

Afin d'améliorer la première lecture de ce nuage de point, nous avons partagé le graphique en quatre cadrans permettant de construire un axe de tendances à la surévaluation ou à la sous évaluation des aptitudes des candidats à suivre la formation. Nous retrouvons les « premiers admis – premiers reçus » dans le quart inférieur gauche du graphique et le « derniers admis – derniers reçus » dans le quart supérieur droit. Cette figure n'est pas destinée à produire un résultat appréciable numériquement dans le sens où il existe une limite trop nette entre chaque cadran. En effet cette représentation tranchée ne tient pas compte des ajustements possibles sur deux années d'études. Ce qui compte réellement est

la tendance que peut avoir un étudiant à conserver une place proche de celle qui lui a été attribuée à l'issue du concours d'entrée.

Nous avons donc réalisé une seconde approche plus réaliste, parce que moins tranchée, qui laisse la place à une zone d'incertitude. Cette zone centrale est calculé sur la base d'un « petit » tiers (ou d'un « gros » quart) de bons à très bons étudiants comme limite supérieure de la zone. La limite inférieure est calculée de la même façon pour les moins bons étudiants. Nous précisons que ce « gros » quart représente précisément 27%, chiffre communément admis dans les approches docimologiques destinées à mettre en évidence les variations des extrêmes⁷³. Il en résulte une partie centrale de quarante six pour cent qui représente la zone d'incertitude d'évolution des étudiants durant leur parcours de formation. La figure ci-dessous illustre cette zone d'incertitude d'évolution sous la forme d'une zone grisée.

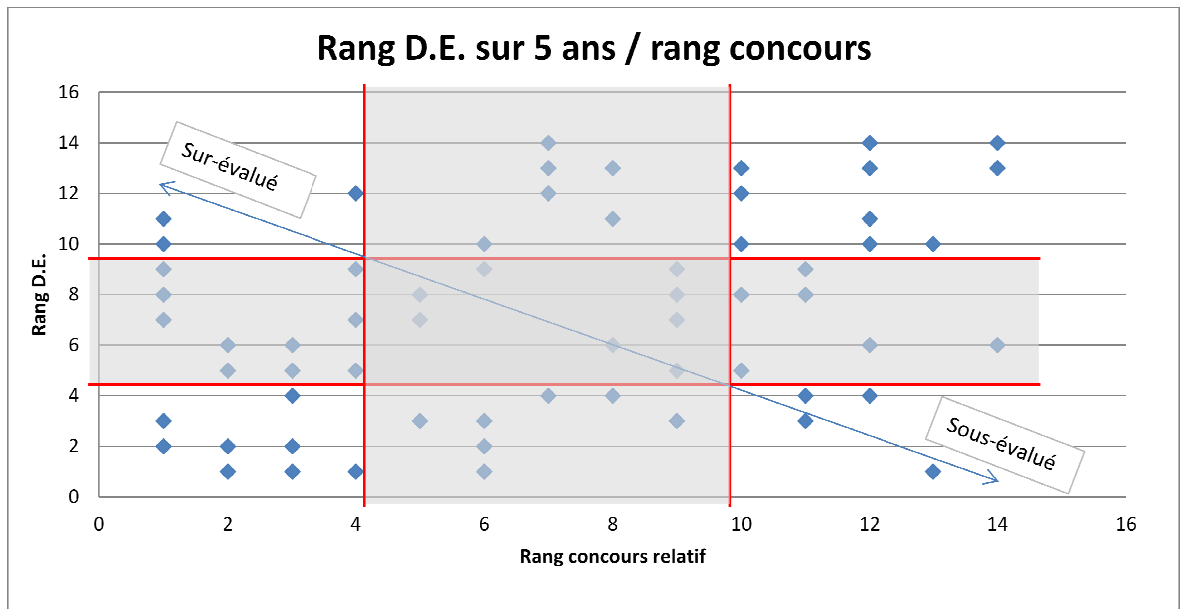


Figure 6: Rang au Diplôme d'Etat / Rang relatif au concours et zone d'incertitude

Cette fois ci, la représentation graphique est plus utilisable. Nous pouvons remarquer la distribution dans les quartiers les plus extrêmes du nuage de points. Les individus situés dans les espaces de sur et sous évaluation sont maintenant plus identifiables. Leur nombre est de sept sur les soixante huit cas étudiés. Les étudiants dont la position n'a pas variée (quart supérieur droit et inférieur gauche) représentent maintenant dix neuf personnes sur soixante huit. Nous pouvons donc avancer qu'il existe environ dix pour cent d'étudiants (10,1% précisément) pour lesquels l'évaluation initiale n'est pas confirmée en fin de

⁷³ Findley W. G. (1956). « A rationale for evaluation of item discrimination statistics » In. Educational and Psychological Measurement,

formation. Ce chiffre est encourageant et peut s'expliquer par le nombre important d'examens (dix) au cours des deux années du cursus de formation et par tous les aléas inhérents aux sujets traités, à leurs corrections et à l'état physique et psychologique de l'étudiant lors de ces épreuves.

La seconde constatation concerne les zones inférieures gauches et supérieures droites représentant respectivement les "bons" et les « moins bons » confirmés dans leur classement initial. Ceux-ci représentent près de vingt sept pour cent de l'échantillon considéré.

Ces résultats laissent à penser, à priori, que l'évaluation des aptitudes à suivre la formation lors du concours d'entrée serait suffisamment pertinente et fiable. Cette première approche graphique des données nécessite d'être confirmée ou infirmée par une étude statistique.

2.4. Analyse statistique et correction méthodologique

A la suite de cette analyse graphique nous pouvons poser l'hypothèse qu'il existe une corrélation entre le rang attribué au concours et le rang à l'issue de la formation. Afin de tester cette hypothèse nous avons construit la droite de régression et calculé le coefficient de corrélation à l'aide du logiciel EXCEL[®] (voir graphique ci-dessous)

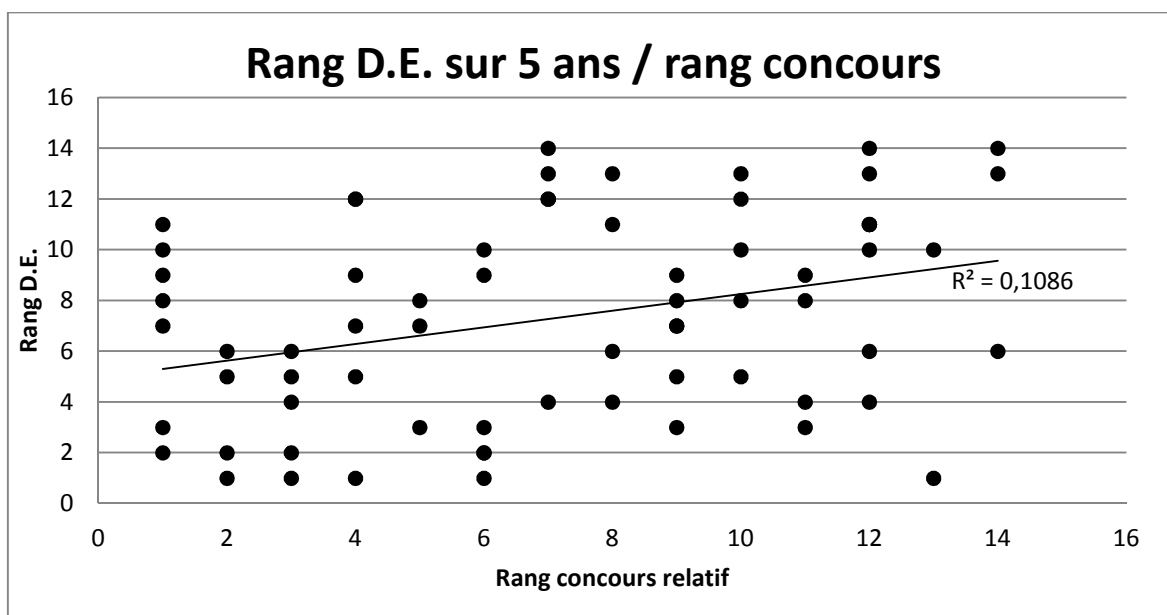


Figure 7: droite de régression et corrélation

La valeur calculée du coefficient de corrélation est de 0,108. Pour mémoire, nous rappelons que ce coefficient de corrélation (ou coefficient de détermination selon les appellations) est compris entre 0 et 1. Plus sa valeur est proche de 1 plus la relation entre les deux éléments est probable ; cela signifie également que plus la valeur de R^2 est élevée, plus celle d'une variation non expliquée est petite. Il est commun d'admettre qu'une corrélation supérieure à 0,7 a peu de chance d'être fortuite. En conséquence cette valeur particulièrement basse infirme très nettement l'hypothèse de départ au point de la rejeter sans autre forme de procès. Les constatations observées lors de notre précédente phase exploratoire nous amènent à nous interroger sur la validité du test utilisé. En effet, une corrélation faible peut être envisageable d'un point de vue statistique mais pas au point de rejeter l'hypothèse de façon aussi catégorique. Nous avons donc recherché le type précis du calcul effectué dans ce cas par le logiciel EXCEL[®]. Il s'avère que ce logiciel propose de manière systématique le calcul du coefficient R^2 de PEARSON sans possibilité d'alternative à notre connaissance. La particularité du coefficient R^2 est de travailler sur des échantillons de données réparties de façon gaussienne. Dans le cas qui nous intéresse, les variables considérées ne sont pas du type « variables continues » réparties selon une courbe en cloche, mais des variables discrètes de type ordinales. Le test statistique à utiliser dans ce cas est donc plutôt le test de corrélation de SPEARMAN adapté aux séries de données non paramétriques de type ordinales⁷⁴.

Suite à ce constat, nous avons donc procédé à une correction méthodologique en utilisant le logiciel PRISM[®] spécialement dédié aux analyses statistiques et permettant le calcul du coefficient de corrélation de SPEARMAN. L'utilisation de ce test est plus délicate que celui de PEARSON dans le sens où il est nécessaire d'utiliser une table spécifique tenant compte du nombre de degrés de liberté considéré en fonction de l'intervalle de confiance choisi⁷⁵. En ce qui concerne l'intervalle de confiance, la valeur communément admise est de quatre vingt quinze pour cent, ceci correspond à une valeur de probabilité inférieure à cinq pour cent que le lien soit dû au hasard, exprimée sous la forme : $p < 0,05$ (ou $\alpha=0,05$ au maximum selon les terminologies employées). Le degré de liberté est conditionné par le nombre d'observations. Pour notre étude, celui-ci est égal à soixante huit paires de données observées soit soixante huit degrés de libertés.

⁷⁴ Mion G. (2009). « *Eléments indispensables de statistiques médicales. Partie 4 - Méthodes non paramétriques* » p. 24

⁷⁵ Voir annexe 2

Le résultat de ce test statistique (voir le tableau suivant) fait apparaître un coefficient de SPEARMAN de près de 0,33 ce qui signifie une corrélation significative entre le rang à l'entrée en formation et le rang à la sortie de cette même formation. Ce résultat infirme de façon très claire les résultats obtenus préalablement et mettent l'accent sur le biais méthodologique qui aurait pu nous fourvoyer dans une analyse erronée du problème étudié.

Number of XY Pairs	68
Spearman r	0,3297
95% confidence interval	0.09192 to 0.5320
P value (two-tailed)	0,0060
Exact or approximate P value?	Gaussian Approximation
Is the correlation significant? (alpha=0.05)	Yes

Figure 8: résultats du test de SPEARMAN

L'analyse de ce type de test statistique réalisé par un logiciel dédié, demande une vérification plus poussée. L'extrait de la table de SPEARMAN présenté ci-dessous fournit les valeurs critiques au-delà desquelles les coefficients de corrélation de SPEARMAN obtenus sont significatifs. Deux possibilités sont proposées : soit un intervalle de confiance de quatre vingt quinze pourcent ($\alpha = 0.05$), le plus communément admis, soit un intervalle de quatre vingt dix neuf pour cent ($\alpha = 0.01$), beaucoup plus rare. La table ne nous permet pas une finesse d'analyse à soixante huit degrés de liberté. Nous considérerons donc le moindre degré de liberté correspondant à soixante pour assurer dans un premier temps le niveau de fiabilité du test.

n	$\alpha = 0.05$	$\alpha = 0.01$
50	0.279	0.363
60	.0255	0.331
70	0.235	0.307
80	0.220	0.287

Figure 9: extrait de la table de SPEARMAN

Si nous considérons que le nombre d'observations est plus proche de soixante dix, le résultat obtenu nous permet de conclure que de toutes façons, il existe une corrélation certaine entre les rangs à l'entrée et à la sortie de la formation dans un intervalle de confiance proche de quatre vingt dix neuf pour cent.

2.5.En conclusion : sommes nous fiables en terme d'évaluation initiale ?

En conclusion de ce chapitre nous pouvons affirmer, au regard de l'étude évoquée ci-dessus, que la méthode globale de sélection des postulants à l'entrée à l'école d'Infirmiers Anesthésiste est fiable. Il reste à démontrer que toutes les étapes de cette sélection sont indemnes de critiques. Nous pouvons en effet considérer qu'un effet Rosenthal puisse être pris en considération et biaiser le résultat obtenu. Quoi qu'il en soit, le résultat est là et les établissements faisant appel à nos services semblent satisfaits des professionnels formés, dans le sens où ils continuent à nous envoyer des candidats, alors que d'autres établissements plus proches de chez eux proposent à priori les mêmes prestations.

Le constat global particulièrement positif que nous venons de faire ne doit pas occulter la possibilité d'une recherche d'amélioration de la méthode de sélection que nous utilisons actuellement.

Le chapitre suivant nous permettra de préciser la notion de sélection dans le cadre d'un examen classant et ce que nous entendons sous le vocable de sélection.

Dans ce contexte, le concept de discrimination qui sera largement utilisé dans la troisième partie de notre travail doit être étudié de façon plus précise.

3. L'EXAMEN ET LES QUESTIONS QUI LE COMPOSE : LA NOTION DE SELECTION ET LE CONCEPT ASSOCIE DE DISCRIMINATION

Les examens et les examens classants (ou les concours) ont en commun la fonction d'évaluation. Ce qui les différencie le plus est la fonction de sélection, spécifique des examens classants et des concours. Ces derniers ont en effet pour objectif d'attribuer un rang de classement aux candidats dont l'évaluation est jugée suffisante. Par la suite, seuls les premiers classés sont retenus pour figurer sur la liste principale d'admission. Ce type de fonctionnement consiste donc en une sélection effectuée sur la base d'un tri préalable en fonction de critères prédéterminés recherchés lors de l'épreuve.

Les premiers auteurs qui se sont penchés sur les qualités requises de ce type de procédure, afin d'obtenir une sélection plus fine des candidats ont, dès le départ, utilisé le terme de discrimination pour évoquer le pouvoir, ou l'effet discriminant d'une question. En premier

lieu, FINDLEY en 1956⁷⁶ proposait son « Indice D », dit indice de discrimination, propre à quantifier numériquement l'aptitude d'une question à obtenir une réponse de bonne qualité dans le groupe des meilleurs d'une épreuve écrite. Par la suite, en 1972, GROOT⁷⁷ exprime la formule de l'index de discrimination (IDD de GROOT), puis plus près de nous, le groupe suisse Édumétrie⁷⁸, et d'autres encore, en reprenant les travaux des précurseurs, continuent d'utiliser le vocable de discrimination.

En 2012, le simple fait de parler de discrimination revêt une connotation nettement péjorative liée au terme. Il nous faut donc préciser les différentes définitions attribuées à ce mot et celle que nous désirons retenir.

3.1. De la discrimination

3.1.1. Les définitions générales

L'origine latine du mot discrimination provient du verbe « *dicriminare* » qui signifie séparer ou trier et de son substantif « *discriminatio* » séparation.

Le dictionnaire Larousse donne deux définitions principales différentes. La première est la suivante :

« Action de séparer, de distinguer deux ou plusieurs êtres ou choses à partir de certains critères ou caractères distinctifs ; distinction : Opérer la discrimination entre l'indispensable et le souhaitable. »

Comme nous pouvons le constater, la signification étymologie originale est conservée et ne revêt aucun sens péjoratif. Le synonyme « distinction », proposé par les auteurs, confirme ce constat en suggérant un terme plus neutre.

La seconde définition montre un tout autre visage qui tient compte de la notion de discrimination sociale :

« Fait de distinguer et de traiter différemment (le plus souvent plus mal) quelqu'un ou un groupe par rapport au reste de la collectivité ou par rapport à une autre personne : Le sexisme est une discrimination à l'égard des femmes. Discrimination raciale. »

⁷⁶ Findley W. G. (1956). « A rationale for evaluation of item discrimination statistics »

⁷⁷ Groot G. (1972) : « Le problème de l'évaluation dans l'enseignement des langues vivantes »

⁷⁸ Cardinet J., et Tourneur Y. (1985) : « Assurer la mesure »

La notion de discrimination sociale apparaît vers la moitié du vingtième siècle lors des grands mouvements égalitaires un peu partout dans le monde. Ceux-ci couvrent les champs des inégalités entre les femmes et les hommes, entre les races ou les minorités, les salaires etc La différence de traitement constatée dans ces différentes situations donne toute la dimension péjorative au terme discrimination que nous lui connaissons actuellement. Les instances politiques et par suite juridiques, ont tenu compte de cette évolution des représentations en considérant qu'il existait un délit pénal de discrimination, en voici la traduction exprimée par l'article 225-1 du code pénal :

« Constitue une discrimination toute distinction opérée entre les personnes physiques à raison de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, de leur patronyme, de leur état de santé, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une nation, une race ou une religion déterminée. »⁷⁹

Nous constatons que la notion de discrimination sociale ressort de manière évidente et officialise ainsi non seulement le côté péjoratif, mais surtout le caractère délictuel du mot qui identifie maintenant des faits de société forts préjudiciables et répréhensibles.

3.1.2. Des confusions qui renforcent l'aspect péjoratif

Le poids des représentations historiques, renforcé par l'appropriation du mot par la justice occulte le sens initial toujours en usage. L'emploi de ce concept est de plus souvent associé à celui de ségrégation ce qui rajoute à la confusion.

En effet le mot ségrégation est issu du latin « segregare » qui signifie mettre à l'écart. Cette notion de mise à l'écart nécessite, certes un tri préalable, mais sa portée est tout autre. Ce concept sociologique retient en particulier une notion d'intentionnalité plus ou moins volontaire de mise à l'écart de groupes raciaux, sociaux ou linguistiques (Grafmeyer1994⁸⁰). Sur le plan politique, le concept trouve ses ancrages sur au niveau racial ou social, sous forme de lois idéologiques promulguées par des gouvernements ségrégationnistes à l'encontre de groupes ethniques (tel l'apartheid en Afrique du sud) ou sociaux (homosexuels dans l'Allemagne nazie). Il existe également une ségrégation

⁷⁹ Article 225-1 du *Code pénal*

⁸⁰ Yves Grafmeyer, 1994, *Sociologie urbaine*, Nathan, Paris

spatiale, souvent liée aux notions de race ou d'appartenance à des groupes sociaux financièrement défavorisés, et associée à une notion de déplacement ou de concentration géographique conduisant à la mise en place de ghettos⁸¹. A ce sujet nous remarquerons qu'il existe également une ségrégation spatiale financière qui conduit aux mêmes conséquences. A noter l'emploi du vocable ségrégation en sciences physiques, évoquant un mélange imparfait dont les éléments n'échangent pas entre eux (sur le plan physique ou chimique). Cette dernière utilisation du terme illustre bien le sens social politique. Si une ségrégation sous entend toujours une discrimination préalable, l'inverse n'est pas systématiquement vrai.

Une autre notion participe à une représentation négative du mot, celle de discrimination positive. En effet, celle-ci sous entend que si elle n'est pas positive, une discrimination ne peut être que négative. Cette notion est apparue aux Etats-Unis par l'intermédiaire de J. F Kennedy sous sa forme originale d' « affirmative action ». La traduction qui en a été faite est en partie responsable de la l'interprétation qui en découle. Cette expression peut d'ailleurs être considérée comme un oxymore⁸² lorsqu'elle est employée dans un contexte politique⁸³.

Le dictionnaire Littré nous fait remarquer de son côté que le terme de discrimination est:

« Terme de psychologie employé par Bain⁸⁴. Faculté de discerner, de distinguer. »

Cette précision nous resitue le concept dans le champ des sciences humaines et en particulier dans son acception psychologique. Celle-ci rejoint la définition originale du mot, sans aucun sens péjoratif. Nous rappelons que les docimologues étaient essentiellement des psychologues, ce qui peut expliquer l'utilisation de ce terme par W. G. FINDLEY et ensuite par ses successeurs lors de leurs études sur les examens classants.

⁸¹ A ce titre, nous pouvons faire remarquer que le terme de ghetto renvoi au ghetto juif de Venise qui était simplement un quartier fermé de la ville sans aucune connotation péjorative de mauvais traitement ou d'insalubrité. L'isolation voulue à cette époque par la communauté juive ne posait aucun problème. Le vocable a été repris par la suite pour évoquer une mise à l'écart de population dont on ne voulait pas reconnaître la différence et dont l'éloignement ou la mise à l'écart satisfaisait les idéologies politiques.

⁸² Combinaison dans un même groupe syntaxique de deux notions contradictoires. Ayant pour objectif d'interpeler le lecteur ou l'auditoire.

⁸³ Nous remarquerons qu'une discrimination positive poussée à l'extrême tendrait à éliminer des postulants particulièrement compétents au profit d'autres candidats afin de respecter une politique de quotas par trop excessive.

⁸⁴ Alexander Bain (1855) « les sens et l'intelligence »

3.1.3. Sélection ou discrimination ?

En conclusion nous pouvons retenir que le terme de sélection est approprié pour exprimer la finalité d'un examen classant, de nature à retenir les candidats dont l'évaluation générale est la meilleure. Celui de discrimination se rapporte plus à la propriété d'une question inscrite dans le cadre de l'épreuve de sélection au regard des qualités intrinsèques qui lui confèrent son pouvoir discriminant

Nous pouvons maintenant aborder l'approche docimologique de la sélection par la représentation graphique édumétrique qui en est faite.

3.2. La courbe de GAUSS : un culte, pas une fatalité

L'étude des résultats chiffrés obtenus par les étudiants ou les candidats est pratiquement toujours présentée par une courbe qui représente le nombre de fois où une note (ou plutôt une tranche de notes) est attribuée. Cette technique est appelée étude de répartition de fréquence, en référence à la fréquence à laquelle apparaissent les notes situées dans un même intervalle. Si l'on se réfère à la Loi de POSTHUMUS⁸⁵, les correcteurs ont tous une tendance personnelle à situer leurs évaluations chiffrées autour d'une moyenne qui leur est propre. De fait, la constatation récurrente du phénomène lui donne raison, d'autant plus que le nombre d'observations est élevé. L'allure de cette courbe est celle d'une cloche plus ou moins « pointue » représentative de la loi de GAUSS et LAPLACE. Les propriétés principales de cette distribution également appelée « distribution normale » sont de présenter une moyenne, une médiane et un mode⁸⁶ très proches voire confondus. L'illustration suivante présente l'allure caractéristique de ce type de courbe : l'axe des abscisses représente les notes (ou les intervalles de notes), l'axe des ordonnées celui de leur fréquences d'apparition (le nombre de notes comprises dans l'intervalle).

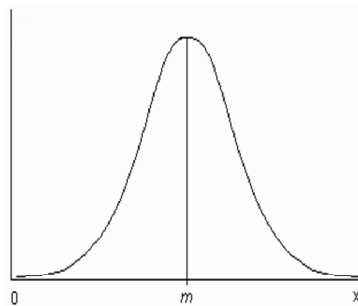


Figure 10: allure caractéristique d'une courbe de GAUSS

⁸⁵ Op sit.

⁸⁶ Valeur de la variable dont la fréquence est maximale.

Cette loi est couramment observée lors de phénomènes naturels étudiés en sciences physiques, expérimentales et naturelles. Le fait qu'il en soit fréquemment de même dans les études docimologiques contribue à considérer cette courbe comme un culte⁸⁷ par la communauté des enseignants, des formateurs ou de toutes personnes amenées à attribuer des notes lors d'une évaluation.

Nous pouvons remarquer, comme cela a été exposé dans la première partie du travail, que ces résultats tiennent essentiellement au comportement de l'examineur lors de sa correction et à la nature ou à la forme de la question qu'il a lui-même élaborée. En accord avec D. LECLERCQ et Coll., nous pensons qu'il est possible et souhaitable dans certains cas de sortir « *de la fatalité de la courbe de GAUSS*⁸⁸ », en particulier lors des examens classants. En effet dans ce cas, l'objectif de l'épreuve est bien de sélectionner les candidats aptes à suivre la formation de ceux qui ne le sont pas, c'est-à-dire de créer deux groupes distincts, idéalement distribués de part et d'autre de la moyenne théorique de l'examen.

Pour illustrer ce propos, nous considérerons une population de candidats à priori théoriquement homogène en termes de connaissances et de compétences au regard des pré-requis à l'inscription (figure 11). La courbe est donc de type Gaussienne régulièrement répartie autour d'une moyenne (m) qui lui est propre.

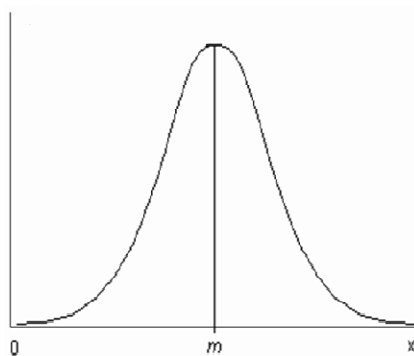


Figure 11: Distribution théorique des candidats au concours

A l'issue de l'examen, l'objectif est d'obtenir une courbe bimodale (figure 12) propre à identifier dans le groupe de départ, les individus dont l'évaluation des aptitudes à suivre la formation sont suffisantes, de ceux pour qui elles ne le sont pas.

⁸⁷ Bonboir A. (1972). « *La docimologie* » p. 94.

⁸⁸ Leclercq D., Nicaise J., Demeuse M. (2004) : « *Docimologie critique : Des difficultés à noter les copies et d'attribuer des notes aux élèves* » p 280

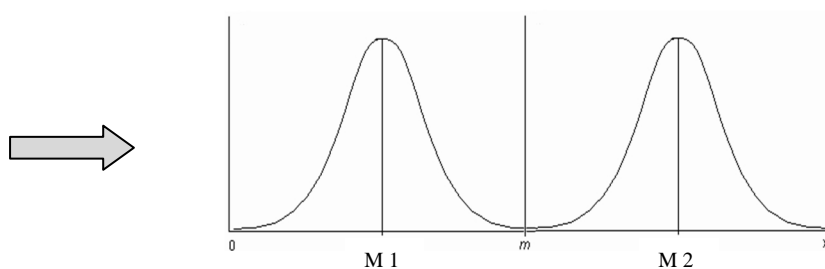


Figure 12: Distribution souhaitée à l'issue du concours

Le mode supérieur (M 2) représente le groupe des meilleurs. Il est situé au dessus de la moyenne théorique d'admissibilité (m) et le mode inférieur (M 1) représentant les moins bons, en dessous de m . Ce type de courbe est caractéristique de deux populations initialement mélangées et mises en évidence par une mesure spécifique des critères désirés⁸⁹.

C'est en cela que les concours diffèrent des évaluations au cours et en fin de formation. En début de formation, les résultats obtenus dans certaines matières spécifiques permettent d'observer des courbes multimodales variables d'un sujet à l'autre. L'objectif de la formation est d'obtenir au final un groupe homogène dont le mode (M 1) est supérieur à la moyenne théorique (m) quel que soit le sujet traité.

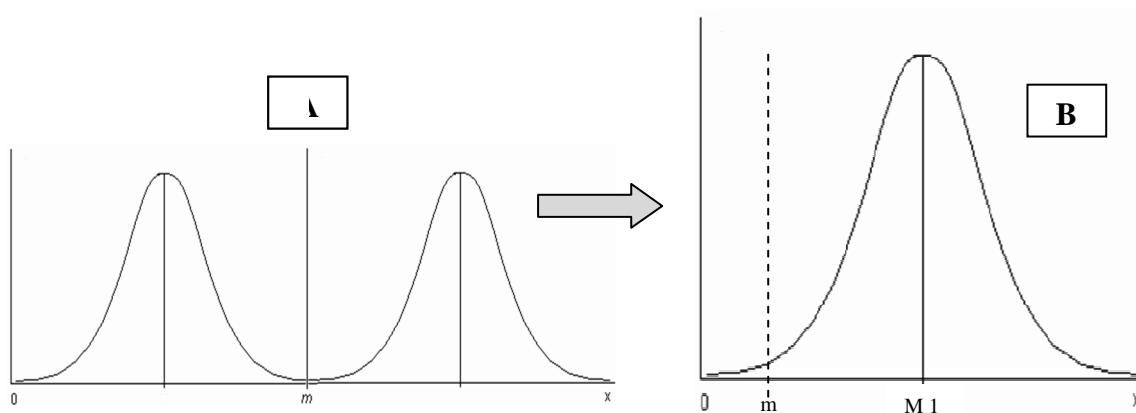


Figure 13: répartition à l'arrivée en formation (A) et répartition souhaitée à la fin des études (B)

L'expression graphique des idées énoncées au cours de ce chapitre nous permet de mesurer l'écart entre un examen et un examen classant. Le point de vue de G. DE LANDSHEERE⁹⁰ sur le sujet résume bien la situation:

« Instruire n'est pas sélectionner. Au contraire ! »⁹¹

⁸⁹ Par exemple : la mesure du poids ou de la taille dans un échantillon comportant des femmes et des hommes.

⁹⁰ De Landsheere G. (1933-2001) titulaire de la chaire de pédagogie expérimentale de l'université de Liège

⁹¹ De Landsheere G. (1971).. « Evaluation continue et examens. Précis de docimologie » p 179

Effectivement l'opposition est totale. Si la formation se doit de tendre vers une unité de compétences et de connaissances, la notion de sélection, spécifique d'un concours, doit tenir compte de la nécessité d'une répartition des candidats en deux groupes distincts : celui dont le mode est supérieur à la moyenne théorique, et donc admissible, et celui des non-admissibles, dont le mode est inférieur à cette limite.

L'objectif de l'examen classant étant maintenant précisément exposé, il nous faut examiner l'épreuve elle-même et ses composantes humaines et techniques.

4. UN CANDIDAT, UN JURY ET DES EPREUVES : LES TROIS COMPOSANTES DE L'EXAMEN D'ENTREE

Parmi les différents auteurs qui ont étudié ce sujet, nous retiendrons principalement P. DAGUE qui identifie en 1952⁹² ce qu'il considère être « *les trois facteurs principaux* ». Ceux-ci sont le candidat, les épreuves et l'examineur⁹³. Jusqu'alors, seuls les deux derniers étaient pris en considération. L'objet de ce chapitre sera d'étudier le rôle et le fonctionnement de ces trois composantes et de leurs éléments constitutifs. Le terme d'examineur sera remplacé par celui de jury puisqu'il s'agit d'une responsabilité collective.

L'objectif de cette étude est de faire ressortir les points susceptibles d'être améliorés dans le cadre du concours qui nous intéresse. Ces éléments de réflexion contribuent à affiner la problématique qui sera traitée dans la troisième partie de notre travail.

4.1. Le candidat

Elément central d'un concours, le candidat est certainement la composante la moins maîtrisable des trois. Ses particularités sont liées à sa personnalité et à ses capacités à répondre aux attentes du jury. Pour lui, les enjeux sont d'ordres psychosociaux liés à l'échec ou à la réussite d'un examen qui lui permettra d'évoluer professionnellement et socialement. Son image se verra affectée positivement ou non en fonction de son résultat à l'épreuve. Il faut tenir compte du fait qu'il s'agit d'un adulte, professionnel reconnu, qui cherche à s'épanouir dans une carrière lui permettant d'être plus autonome dans son travail. D'un point de vue psychologique, nous pouvons le considérer comme un individu

⁹² Dague P. (1952). « *Le régime des examens et des concours en France : problèmes de docimologie* ». In: *Enfance. Numéro Spécial*, 1952. Pp 450 - 457

⁹³ Signalons que F. BACHER (1969) qualifie ces éléments de trois sources d'erreurs

qui tend à évoluer vers la dernière étape de la pyramide de MASLOW. En cas d'échec, il faudra donc savoir lui expliquer que ceci ne remet pas en cause sa valeur professionnelle et qu'un concours reste une épreuve ou la marge d'incertitude est élevée, que ce soit sur les sujets abordés ou le niveau général des candidats présents cette année là.

Il faut bien reconnaître que la position de candidat n'est pas facile à tenir. En effet, celui-ci doit, dans un espace de temps très court, faire la preuve de ses connaissances et de ses compétences. La tâche est ardue et demande beaucoup de résistance physique et psychologique. P. DAGUE fait remarquer : « ... *que l'on préjuge ainsi autant de la résistance physique et de la stabilité émotionnelle que de ses aptitudes intellectuelles.* » C'est tout à fait vrai, d'autant plus dans le cas de candidats à la formation d'IADE, où la gestion du stress en situation de pression émotionnelle est un élément de compétence recherché. Nous pouvons faire remarquer à ce sujet que tous les élèves ou étudiants inscrits dans une formation en France sont, selon la dernière enquête PISA⁹⁴ de 2009, de tous les Européens, les plus stressés à l'idée de devoir subir un examen.

Par ailleurs le candidat ne peut pas faire complètement la preuve de ses connaissances et de ses compétences professionnelles dans le cadre de ces épreuves. Le choix des sujets abordés est certes large, mais ne permet pas une analyse exhaustive des capacités du futur étudiant, ce qui fait dire à P. DAGUE : « *Ce n'est pas l'élève (ou le futur étudiant – Ndr -) que l'on juge, mais le candidat, placé dans des conditions artificielles et peu propices à en donner un portrait fidèle* ».

Nous ferons remarquer que, toujours selon le même auteur, si « ... *Le candidat est toujours jugé en fonction de l'échantillon auquel il appartient* », ceci reste une des particularités d'un examen classant et que, contrairement à un examen classique, l'effet de comparaison est explicitement recherché.

Il reste à considérer l'influence d'une préparation au concours. L'une des missions des écoles d'IADE est de mettre en place un dispositif de formation qui vise à préparer les futurs candidats aux épreuves de sélection. Il est bien entendu que notre rôle reste celui d'un organisateur et qu'en aucune façon nous ne prenons une part active à cette prestation. Ceci nous placerait dans la position à la fois de juge et de partie, ce qui d'un point de vue éthique serait inacceptable. Le constat que nous pouvons faire est que, dans la mesure où nous n'avons pas toujours connaissance des antécédents de candidats, la grande majorité des postulants admis en tête de liste ont bénéficié d'une préformation à hauteur d'environ

⁹⁴ Programme for International Student Assessment

quatre vingt pour cent. Pour les ajournés et les derniers classés, la proportion s'inverse. Ceci nous porte à croire que ce type de pratique est efficace; néanmoins ceci a un coût et même si celui-ci est le plus souvent supporté par les établissements employeurs, ce n'est pas toujours le cas. Il existe donc une inégalité liée à cette pratique qui hypothèque les chances de réussites.

En conclusion nous dirons que la composante « candidat » laisse peu de potentiel d'action dans le cadre d'une amélioration de l'efficacité des épreuves d'un examen classant. Nous retiendrons malgré tout la possibilité de généraliser la pratique des sessions de préformations aux concours d'entrée dans les écoles d'IADE afin de préserver l'égalité des chances des candidats.

Nous pouvons maintenant nous intéresser à « l'autre coté de la table », à savoir le jury qui doit évaluer les qualités du candidat à suivre la formation.

4.2.Le jury

Le jury de l'examen d'entrée à l'école d'IADE a la particularité de présenter une unité professionnelle. En effet, chacun d'entre nous appartient à la communauté de l'anesthésie. Au sein de ce groupe il existe une diversité hiérarchique mélangeant un Professeur des Universités titulaire de la chaire d'anesthésie, des Maitres de Conférence Universitaires en anesthésie, des Médecins Praticiens Hospitaliers spécialisés en anesthésie et réanimation, le Directeur de l'école d'anesthésie, le Cadre Infirmiers Anesthésiste enseignant formateur de l'établissement, des Cadres IADE et des Infirmiers Anesthésistes issus du terrain. L'objectif de toutes ces personnes est de sélectionner des candidats susceptibles de suivre la formation d'Infirmier(e) Anesthésiste. L'idée d'une profession porteuse de valeurs fortes est présente en chacun de nous. La notion de corporatisme n'est pas absente de cette union. Nous pouvons dire que nous nous inscrivons dans la lignée d'un apprentissage de type compagnonnique qui ne dit pas son nom, mais n'est ce pas coutumier aussi dans d'autres professions bien organisées.

Nous allons maintenant examiner les différentes influences des personnalités qui composent le jury et préciser le rôle de celui-ci.

4.2.1. Les forces en présence

Dans un groupe de ce type, composé de huit personnes aux statuts hiérarchiques variés, nous pouvons nous interroger sur l'influence de chacun d'entre nous sur le fonctionnement du « système jury ». Pour H. MINTZBERG, influence et pouvoir sont deux concepts qui peuvent être confondus⁹⁵

CROZIER et FRIEDBERG (1977) ont défini ce qu'ils considèrent être les quatre sources du pouvoir. Celles-ci sont toutes liées à la maîtrise des zones d'incertitudes : « *ce qui est incertitude du point de vue des problèmes est pouvoir du point de vue des acteurs* »⁹⁶.

- La première source est la possession d'une compétence ou d'une spécialisation. Elle est caractérisée par un monopole de l'expertise dans le domaine considéré.
- La seconde source concerne les relations entre l'organisation et son environnement. C'est le pouvoir du « *marginal sécant* » c'est à dire un « *acteur partie prenante de plusieurs systèmes en relation les uns avec les autres* ».
- La troisième source est le contrôle de la communication interne. Celle-ci tient à une connaissance large et actualisée de l'information.
- Enfin, la quatrième source consiste en une connaissance des règles organisationnelles et de leurs utilisations. Celle-ci permet de contourner les règles établies, à priori destinées à limiter les zones d'incertitudes, de façon à en créer d'autres, illégitimes, propres à offrir un levier de pouvoir par le biais d'un possible retour à la règle stricte.

Max WEBER⁹⁷(1962) précise les trois types de légitimité qui sont les fondements de l'autorité :

- « *La légitimité charismatique* » qui est une notion subjective reposant sur les qualités présumées exceptionnelles d'une personne.
- « *La légitimité traditionnelle* » acquise à travers l'ordre établi traditionnellement. Sorte de sacralisation du titre, impossible à remettre en question. C'est une légitimité vécue de façon personnelle lorsqu'on s'insère dans un milieu défini.
- « *La légitimité rationnelle* », assise sur la représentation légale de l'individu. Celle-ci accorde une autorité de principe réglementaire, c'est le cas de la Directrice de notre établissement désignée présidente du jury par l'arrêté qui régit le fonctionnement des écoles d'IADE.

⁹⁵ Mintzberg H. (1986). « *Le pouvoir dans les organisations* » ; Paris : Editions des organisations.

⁹⁶ Crozier, M., Friedberg, E. (1977). « *L'acteur et le système* » p. 24

⁹⁷ Weber M. (1995). « *Économie et société* », Paris. Éd. Plon, Agora. (1re édition 1921)

Nous remarquerons que le Directeur Scientifique, obligatoirement Professeur des Universités qualifié en Anesthésie et Réanimation, tire son pouvoir de ses connaissances dans la spécialité. Il n'est pas le supérieur hiérarchique de la Directrice, mais c'est la légitimité traditionnelle qui lui confère une grande autorité au sein du jury.

Les autres membres puisent les sources de leur pouvoir, pour certains dans leurs connaissances approfondies de l'anesthésie et pour d'autres dans leur rôle de marginal sécant. Chacun peut bien sûr avoir une influence qui découle de plusieurs formes de légitimités.

Ces différentes positions statutaires sont naturelles lorsqu'elles sont établies depuis longtemps, mais elles doivent être travaillées pour permettre à chacun de s'exprimer, en particulier lorsqu'il s'agit d'un membre qui vient d'être sollicité. Pour cela, il va devoir se faire reconnaître comme un interlocuteur valable par l'ensemble du groupe. N. ROELEN⁹⁸ nous apprend que cela tient compte de la valeur sociale des protagonistes. Cette valeur à l'instar des valeurs boursières peut varier en fonction du contexte, il faut obtenir ce que N. ROELEN appelle « une autorisation socio-subjective à soutenir des prétentions » :

«... ne dépend pas essentiellement de leur qualité argumentative ou de leur pertinence, mais du poids du locuteur c'est-à-dire du degré d'autorisation socio-subjective dont il dispose pour élever des prétentions à soutenir son point de vue.⁹⁹ »

A travers l'expression « poids du locuteur », nous pouvons penser que cette affirmation de soi dans la prise de parole écoutée par les autres membres du groupe s'apparente à une forme d'autorité et de prise de pouvoir. Si celle-ci n'est pas reconnue comme légitime par un des interlocuteurs en position de supériorité hiérarchique, il peut s'en suivre une situation conflictuelle. Dans ce cas, le recours au rapport de force est envisageable mais exceptionnel et non souhaitable pour les deux parties, en effet :

« Quand une contradiction entre des points de vues divergents crée un conflit intra et inter-subjectif, les interlocuteurs ont le choix entre s'y confronter ou le supprimer, en instaurant entre eux un rapport de pouvoir qui donnera avantage à un point de vue, par la domination d'un interlocuteur sur l'autre.¹⁰⁰ »

⁹⁸ Roelens N. (2003). « Interactions humaines et rapports de force entre subjectivités » p. 210

⁹⁹ Ibid. p. 211

¹⁰⁰ Roelens N. (2003). « Interactions humaines et rapports de force entre subjectivités » p. 220

Dans ce cas, le dominé, celui qui se résigne à cette situation, peut se mettre en situation d'irresponsabilité et considérer par la suite qu'un dysfonctionnement majeur existe, qu'il n'est pas de son fait, mais de l'entière responsabilité du dominant. Le levier de pouvoir risque de s'inverser et rétablir un équilibre du poids des points de vue. Cela peut être un risque à courir par le dominant, mais il s'agit d'un pari qui peut lui faire perdre beaucoup en terme de position influente et générer d'importants dysfonctionnements au sein du groupe.

En effet, la responsabilité du jury en tant qu'entité, ne tient que par le partage de cette responsabilité entre tous les individus du groupe. C'est une des composantes qui permettent de sceller une cohésion de l'ensemble. Ceci nous conduit à examiner les différents plans qui composent la notion de responsabilité du jury.

4.2.2. La responsabilité du jury

Cette responsabilité s'établit à plusieurs niveaux : juridique, éthique et moral.

D'un point de vue juridique, il s'agit d'une responsabilité partagée entre le jury, l'établissement employeur, et les différentes tutelles, qu'elles soient réglementaires (ARS¹⁰¹ et DRJSCS¹⁰²), financière (le Conseil Régional) ou administrative (Le Centre Hospitalier Universitaire de référence).

Devant la diversité d'un tel paysage, il est facile d'imaginer à quel point la notion de responsabilité partagée peut être diluée, voire ignorée. En effet chacun des acteurs rejette sur l'autre ce qu'il considère comme ne faisant pas partie de ses attributions. Il en résulte un discours entre les tutelles du type « *il*¹⁰³ est de la responsabilité de quelqu'un d'autre » ou du « *on* devrait plutôt se préoccuper de ... » ne permettant en aucun cas d'identifier les responsabilités individuelles ou partagées. L'emploi du « *tu* » ou du « vous êtes responsable » n'est que très rarement mentionné au niveau officiel, mais trouve toute la mesure de son expression dans les « entre-deux » avec des niveaux hiérarchiques inférieurs. Le « *je* », est quant-à lui, plus employé à des fins plus tranchantes, en termes d'injonctions, que responsables d'un point de vue purement réglementaire. Il nous semble donc qu'il s'agit, dans ce cas, plus d'une responsabilité collective de type politique, en

¹⁰¹ Agence Régionale de Santé

¹⁰² Direction Régionale de la Jeunesse, des Sports et de la Cohésion Sociale

¹⁰³ Les termes de « Je, tu, il ils ou on » sont utilisés dans ce cas comme définis par M. boutinet dans son intervention du séminaire transversal.

recherche de preuves et d'imputabilité¹⁰⁴, que d'une réelle responsabilité collective. En effet la responsabilité ne peut pas se résumer à la seule recherche d'une culpabilité. Pour reprendre Max Weber :

« ... le partisan de l'éthique de responsabilité comptera justement avec les défaillances connues de l'homme (...) et il estimera ne pas pouvoir se décharger sur les autres des conséquences de sa propre action pour autant qu'il aura pu les prévoir¹⁰⁵. »

D'un point de vue éthique, la situation est, pour nous, beaucoup plus claire. Un jury se trouve confronté à deux situations : responsable vis-à-vis de qui, et vis à vis de quoi ? En ce qui concerne le premier point, la réponse nous semble aller de soi : il s'agit de l'étudiant envers qui le formateur doit adopter le statut premier de « sponsor » au sens classique du terme¹⁰⁶. Loin de nous l'idée de considérer l'étudiant comme irresponsable. C'est un adulte responsable de ses actes, en formation comme ailleurs. Se pose ensuite la question de vis-à-vis de quoi ? Vis-à-vis des commanditaires cités plus haut, qui représentent notre hiérarchie et les autorités de tutelles ? Oui bien sûr, mais pas seulement. N'oublions pas que la profession est pour nous le principal commanditaire de notre fonction. Nous avons tous été à un moment de notre vie, ou le sommes encore, des professionnels de l'anesthésie et nous nous devons de transmettre notre savoir et nos valeurs professionnelles à ceux qui assureront la continuité du service auprès des patients. A ce sujet, le jury, qui est soumis aux contraintes du passé et de l'avenir d'une profession, nous interpelle¹⁰⁷. Il ne faut pas oublier de permettre aux nouveaux d'apporter ce sang neuf qui fera l'avenir de la profession face à certaines réticences du métier et de ses représentants, conscients de leurs connaissances, de leur savoir faire ainsi que du bien fondé de leurs actions, qui prévaut depuis des années.

C'est à notre sens l'occasion de promouvoir une dynamique porteuse d'idées nouvelles, tout en conservant la conscience des places de chacun. Il faut accepter la dissymétrie du dialogue en gardant à l'esprit l'égalité intellectuelle de tous. L'idée d'une action fondamentale en direction de la profession doit perdurer.

Enfin, du point de vue moral la responsabilité du jury revêt un intérêt tout particulier. Il ne s'agit pas là de parler de « la Morale » au sens politique, éducationnel ou religieux, mais

¹⁰⁴ Au sens d'E. Kant : « être capable d'être traité comme l'auteur de ses propres actes ».

¹⁰⁵ Max Weber « *Ethique de conviction et éthique de responsabilité* »

¹⁰⁶ Employé dans le sens latin original : « être responsable de l'autre ».

¹⁰⁷ Hanna Arendt « *Double responsabilité dans l'éducation* ».

plutôt de la morale professionnelle, peut-être assez proche de l'éthique, mais plus propre à la vision personnelle plus ou moins consciente que se fait le jury de sa profession initiale. C'est à travers ses valeurs ou ses centres d'intérêts particuliers que le jury gravera consciemment ou non l'empreinte d'un établissement et donnera cette valeur ajoutée, ou cette spécificité typique de l'établissement. Cette forme de responsabilité est bien souvent ignorée mais souvent exprimée dans des termes tels que : « *la patte de..* », « *c'est une marque de fabrique* » ou bien « *ça c'est l'école « intel »* ». Ceci nous incite à croire qu'il existe une reconnaissance implicite par chacun des membres de la responsabilité morale du jury dans l'inconscient collectif.

En résumé, le jury est donc un système où les jeux d'influences sont nombreux et lui confère un équilibre fragile. Sa cohésion de groupe et les valeurs professionnelles qui l'animent sont les conditions d'un fonctionnement permettant d'obtenir un résultat satisfaisant. Nous allons nous intéresser maintenant au rôle de ce groupe dans le déroulement du concours de l'examen d'entrée à l'école d'IADE.

4.2.3. Le rôle du jury

Au-delà du rôle bien connu d'évaluateur, le jury a également la charge de la conception des épreuves qui constituent l'examen d'entrée. Ceci est valable pour les parties écrites et orales.

Lors d'une première réunion préparatoire, les questions proposées par les personnes conviées, font l'objet d'une étude tant sur le fond que sur la forme. Une liste de vingt questions choisies est établie. Chaque item est accompagné d'une grille de correction qui est examinée par le collectif et soumis à l'approbation de chacun. Les correctifs nécessaires sont apportés et le document final, comportant les vingt questions posées à l'épreuve écrite, fait l'objet d'une relecture minutieuse au cours d'une seconde réunion avant d'être soumis aux candidats le jour de l'épreuve.

L'accent est principalement mis sur la pertinence des questions écrites, qui ont pour objectif d'évaluer le niveau de connaissances théoriques et le champ de celles-ci. Il est en effet indispensable de tester les éléments de bases nécessaires à la bonne compréhension des enseignements de la formation IADE, mais sans sortir du programme des études dispensées par les Instituts de Formation en Soins Infirmiers.

Il reste ensuite à tenir compte du degré de difficulté. Il faut savoir l'ajuster pour ne pas rendre l'épreuve d'admissibilité hyper-sélective en ajournant un trop grand nombre de

postulants dès la première partie de l'examen, ou au contraire basculer vers trop de facilité. Cette dernière possibilité enlèverait beaucoup de crédibilité au concours et à l'utilité des deux épreuves. A ce sujet, la pondération des notes à attribuer aux réponses est un élément de régulation, tant de la difficulté que de l'importance accordée à l'item par le jury.

Les sujets de l'épreuve orale sont préparés de la même manière, en gardant à l'esprit que celle-ci cherche à évaluer les capacités de réflexion, de synthèse, d'organisation et les compétences professionnelles. Ce dernier point pose la question de la pertinence de ce type d'épreuve à évaluer la compétence, celle-ci ne pouvant théoriquement se faire qu'en situation réelle de travail. Par ailleurs les entretiens d'admissions sont souvent qualifiés de subjectifs et peu fiables par certains auteurs¹⁰⁸. En effet nous admettons que bon nombre de biais docimologiques sont susceptibles d'influencer l'évaluation du jury, parmi eux, nous retiendrons principalement :

- L'effet Rosenthal (ou effet Pygmalion), à notre sens le plus important dans ce cas,
- L'effet de halo
- L'effet de comparaison (ou effet de contraste), même si la finalité d'un concours est de comparer les candidats, l'ordre de passage peut avoir une influence non négligeable.
- Les biais sociaux¹⁰⁹ enfin, qui figurent parmi les plus insidieux. Nous nous en défendons fréquemment mais nous n'en mesurons pas toujours la portée. Nous signalons tout de même une initiative propre à notre établissement sous la forme d'une stricte parité femmes hommes dans la constitution du jury par réalisée par la directrice de l'établissement. Cette mesure n'est pas exigée par la réglementation mais et nous semble judicieuse dans son esprit.

En résumé, malgré le grand nombre de paramètres à prendre en compte, nous avons pu mettre en évidence dans notre étude présentée au début de cette partie ¹¹⁰ que le résultat est conforme à ce que nous sommes en droit d'en attendre, à savoir une évaluation du niveau des candidats entre eux en conformité avec leur niveau de réussite à la fin de la formation.

Nous allons maintenant pouvoir étudier la pertinence de chacune des épreuves constitutives de l'examen, en terme de pouvoir discriminant propre à effectuer une

¹⁰⁸ Plenchette M. (1992) : « *Les entretiens d'admission : méthode et stratégies* » ; in Soins Formation – Pédagogie – Encadrement n°2 ; 2^{ème} trimestre 1992, pp 47 – 51

¹⁰⁹ Merle P. (2010) : « *L'évaluation en philosophie : quels problèmes ?* » In « Côté-Philo » n° 14 p. 13 - 15

¹¹⁰ Chapitre « les premiers seront les derniers ? » (page 45).

sélection claire des candidats. Le poids de chaque épreuve dans l'évaluation finale sera également apprécié.

4.3. Les épreuves

L'étude docimologique que nous avons réalisée a pour objectifs :

- De mettre en lumière l'importance relative des épreuves écrites et orales dans la décision finale,
- De rechercher l'existence d'une corrélation dans les résultats d'évaluation des deux épreuves,
- D'apprécier le pouvoir discriminant de chacune des épreuves par l'étude de la courbe de répartition de fréquence.

L'échantillon étudié porte au maximum sur les sept dernières années d'exercices. Pour des raisons techniques, les données de l'année 2007 concernant les notes finales et celles de l'épreuve orale ne sont plus disponibles. Ceci explique pourquoi l'analyse des notes des épreuves écrites est effectuée sur sept ans, alors que celles des notes finales et des oraux sont sur une période de six ans. Cela ne change rien à la qualité de l'échantillonnage. En effet, les données sont toujours comparées sur la même période.

Nous rajouterons que le nombre de résultats aux épreuves écrites est nécessairement plus important puisqu'il comprend tous les candidats qui n'ont pu accéder à l'épreuve orale, faute d'une note suffisante.

Les outils d'analyses informatiques utilisés sont le tableur Microsoft EXCEL[®] pour le recueil, le tri des données et les représentations graphiques courantes. Le logiciel PRISM[®], spécialisé dans le traitement et l'analyse statistique des données, a été utilisé pour des calculs spécifiques et la réalisation de leurs représentations graphiques.

Nous commencerons par la présentation des résultats concernant la note finale¹¹¹ afin de rendre compte de l'idée générale de la répartition des notes attribuées lors de la sélection terminale.

¹¹¹ Voir ANNEXE III

4.3.1. La distribution des notes finales

Le graphique suivant représente la distribution des notes finales obtenues lors de l'examen d'entrée. Nous trouvons en abscisses les valeurs centrales des intervalles de notes attribuées sur 80 points. Cet intervalle est de cinq points. L'axe des ordonnées représente le nombre de valeurs comptées dans chacun de ces intervalles.

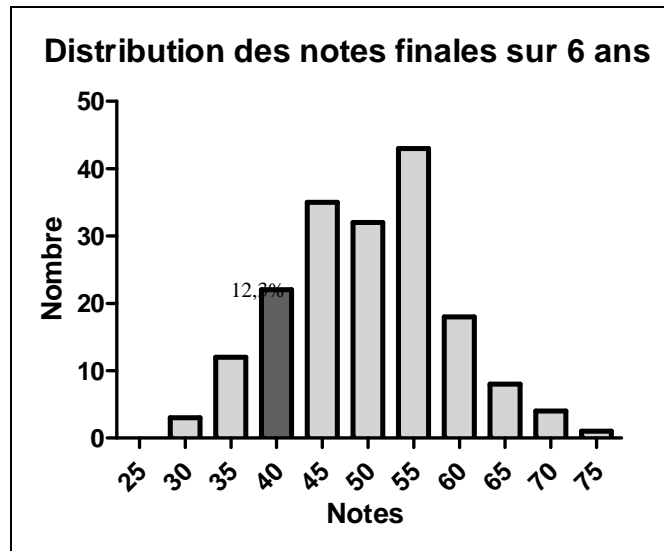


Figure 14: Distribution des notes finales de l'examen

Nous constatons que la courbe obtenue est globalement de type gaussien, si l'on excepte une légère faiblesse sur l'intervalle 47,5 - 52,5, sans conséquences sur l'analyse du segment situé autour de la moyenne qui nous intéresse (à savoir 40 sur 80). Cette représentation met en évidence un mode entre 50 et 55, c'est-à-dire nettement supérieur à la note limite d'admission fixée à 40 sur 80 (moyenne théorique d'admission de l'épreuve). Nous remarquons également qu'il existe un petit peu plus de 12% de candidats dans l'intervalle de 37,5 à 42,5. Ce résultat est acceptable dans l'absolu parce qu'il concerne les candidats derniers classés sur la liste complémentaire. En effet ceux-ci ont peu de chance d'intégrer la formation.

En résumé de la distribution de la note finale :

Le résultat global est donc assez satisfaisant en termes de sélection. Le mode de répartition est très nettement supérieur à la moyenne ce qui signifie que les premiers reçus se sont vus attribuer des notes nettement supérieures à celles des autres candidats. De plus, aucun des candidats sélectionnés sur la liste principale ne risque d'être proche de la limite d'ajournement. En effet, la proportion de reçus sur la liste principale varie selon les années, entre un tiers et la moitié des candidats admis. Une

amélioration serait malgré tout souhaitable pour les notes situées dans les environs proches de la moyenne. Ces constatations peuvent être à mettre sur le compte de deux épreuves dont les résultats chiffrés peuvent se parasiter.

Nous allons maintenant étudier individuellement chacune de ces épreuves. Pour cela, nous analyserons les répartitions des notes, puis nous rechercherons les corrélations possibles de celles-ci avec les notes finales et par la suite, entre-elles. Le coefficient de corrélation obtenu nous permettra de quantifier le poids relatif de l'épreuve étudiée sur l'ensemble du concours.

4.3.2. L'épreuve écrite : une distribution trop gaussienne et peu de poids de dans la décision finale

Chronologiquement, l'épreuve écrite est la première. Ceci implique que l'attribution d'une note inférieure à la moyenne correspond à l'ajournement immédiat du candidat. La qualité discriminante de l'exercice est donc prépondérante pour permettre une sélection claire des candidats. En effet, un résultat proche de la moyenne laisse planer un doute sur les capacités de l'examiné. De plus, il génère chez celui-ci un sentiment de frustration bien compréhensible lorsqu'il est éliminé pour un dixième de point. Nous nous intéressons donc plus particulièrement à l'intervalle de notes centré sur la moyenne. Pour plus de lisibilité les notes sont ramenées à un maximum de 20 points.

Distribution de la note d'écrit :

Lors d'une première approche portant sur un échantillon de 271 personnes, représentant les sept derniers examens d'entrée (voir figure suivante), nous avons retenu à priori un intervalle de deux points pour construire le diagramme de distribution suivant :

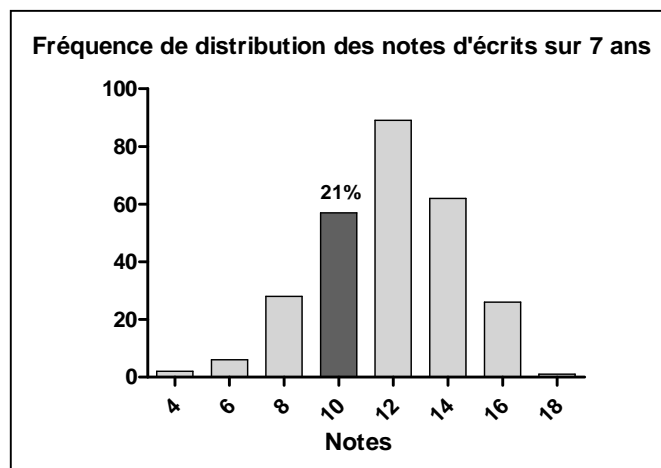


Figure 15: Répartition des notes d'écrits sur 7 ans

Nous observons un nombre important de résultats situés dans l'intervalle allant de 9 sur 20 à 11 sur 20. Celui-ci représente vingt et un pour cent de la population considérée, soit 57 individus sur 271. Ce chiffre très important nous a conduits à reconsidérer la pertinence de l'intervalle choisi. Après réflexion, il nous a semblé plus judicieux de retenir un intervalle d'un point couvrant la plage s'étendant entre 9,5 et 10,5 sur 20.

En effet, si nous reprenons les raisons initiales de ce travail, sur la distribution des notes de l'écrit, nous pouvons considérer que la zone d'incertitude du jury ne s'étend pas à un point d'écart, dans le sens où la note est calculée avec une décimale. Par expérience, il en va de même pour le candidat qui ne met jamais en doute la validité de l'évaluation de sa copie dès lors que la note est inférieure à 9,5 sur 20.

Cela nous permet d'examiner les nouveaux résultats présentés ci- dessous :

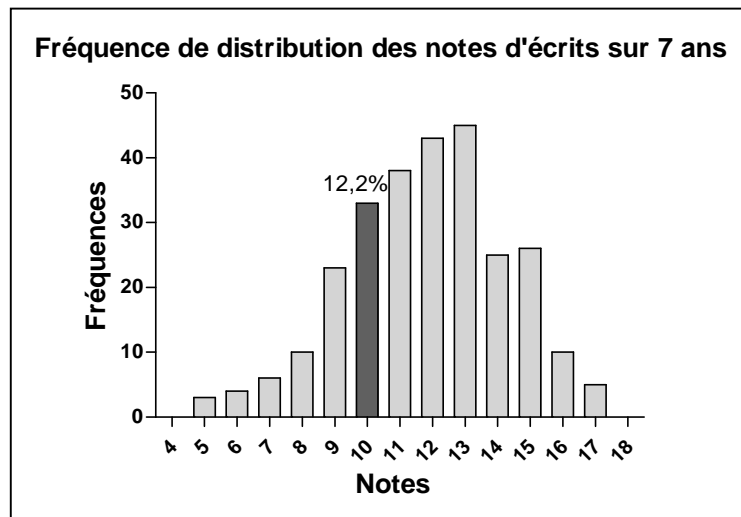


Figure 16: Répartition des notes d'écrits sur 7 ans (Intervalle réduit)¹¹²

En divisant par deux l'intervalle précédemment retenu, nous sommes en droit de penser que le nombre de notes contenues dans cette plage doit être également divisé par deux. Ce n'est pas le cas et cela confirme une concentration des notes autour de dix sur vingt pour l'intervalle considéré. En effet 12,2 % des résultats, soit trente trois candidats sont concernés.

Si nous considérons la répartition gaussienne de l'échantillon¹¹³, le pouvoir discriminant de l'épreuve n'a que peu de valeur. La capacité de sélection permettant d'accéder à l'épreuve orale s'en trouve altérée et pose la question de l'efficacité de l'épreuve au regard d'une

¹¹² Voir Annexe 4

¹¹³ Calcul du test de Shapiro-Wilk : voir annexe 4.

courbe trop gaussienne qui devrait, pour remplir son rôle de sélection, tendre vers une distribution binomiale.

Nous allons maintenant apprécier l'influence (le poids) de cette épreuve dans la décision finale d'admission. Ce travail porte sur six années et non sept comme précédemment car, ainsi que nous l'avons précisé en introduction, les données des oraux et des notes finales ne sont pas accessibles pour l'année 2007¹¹⁴.

Poids de l'écrit dans la note finale :

Afin d'étudier le poids de l'épreuve écrite dans la décision finale, nous avons procédé à une analyse de la corrélation entre les notes de l'écrit et celles du résultat final dont elles sont une des deux composantes. Sur le graphique qui suit, chaque point correspond à un candidat. Pour chacun d'entre eux nous trouvons en abscisses leur note d'écrit et en ordonnées leur note finale. La droite de régression dans ce nuage de points est tracée et accompagnée de son coefficient de corrélation¹¹⁵.

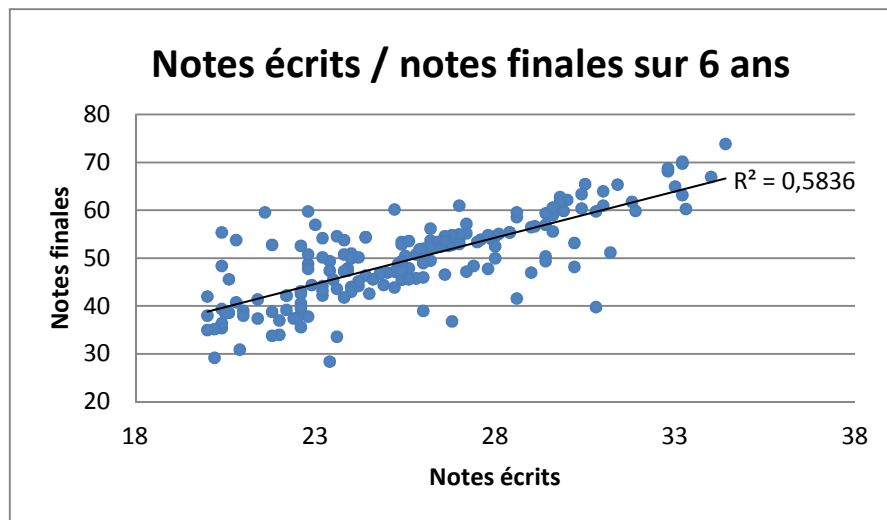


Figure 17: corrélation des notes d'écrits par rapport aux notes finales

Nous remarquons que la droite de régression accuse une pente ascendante, ce qui pourrait laisser penser qu'il existe un lien entre les deux notes. Le calcul du coefficient R^2 montre que ce lien est ténu. En effet, sa valeur est de 0,58, donc inférieure au seuil de 0,70, ce qui ne nous permet pas de dire qu'il existe une corrélation franche entre les deux valeurs. La note d'écrit a donc peu d'incidence sur le résultat de l'évaluation d'admission du candidat.

¹¹⁴ Ibid. annexe 4

¹¹⁵ Le coefficient de corrélation linéaire R^2 , également appelé coefficient de détermination, est dans ce cas le coefficient de Pearson. Celui-ci est compris en 0 et 1 : plus sa valeur est proche de 1 plus la relation entre les deux éléments est probable ; cela signifie également que plus la valeur de R^2 est élevée, plus celle d'une variation non expliquée est petite. Il est commun d'admettre qu'une corrélation de PEARSON supérieure à 0,7 a peu de chance d'être fortuite.

En résumé, l'épreuve d'admissibilité écrite pèse peu dans la décision finale. Nous ajouterons que l'aspect de la courbe de distribution des valeurs qui la composent tend trop vers la Loi Normale, ceci ne lui confère donc pas la qualité discriminante que nous sommes en droit d'attendre d'une épreuve d'un examen classant. Les raisons de ce dysfonctionnement les plus fréquemment évoquées par les auteurs, en particulier J.M. DE KETELE et G. DE LANDSHEERE¹¹⁶ sont, le manque de communication sur les éléments formalisés de correction et l'absence de coordination entre les membres du jury. Dans le contexte qui nous intéresse, nous avons exposé précédemment que ce n'est pas le cas. Le problème doit donc se situer au niveau de la conception des questions constitutives de l'épreuve.

Enfin, il faut rappeler le rôle essentiel de cette étape d'admissibilité à l'entretien d'admission, lequel fait l'objet d'une analyse dans le chapitre suivant.

4.3.3. La distribution et le poids déterminant des notes de l'épreuve orale dans la décision finale

L'échantillon et la méthode utilisés pour cette analyse sont les mêmes qu'au chapitre précédent. Nous commençons donc par la de distribution, qui représente le nombre de candidats ayant obtenu une note comprise dans une intervalle de deux points.

Distribution de la note d'oral

La figure suivante visualise la distribution des notes de l'entretien d'admission sur six ans :

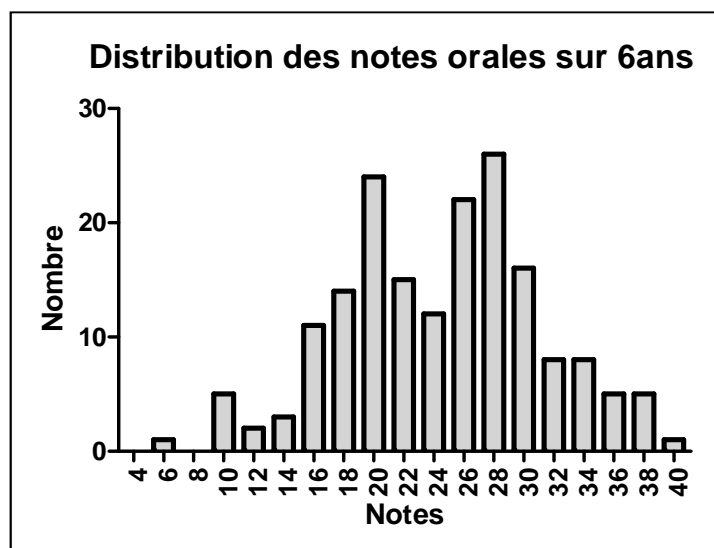


Figure 18: distribution des notes de l'épreuve orale sur 6 ans

¹¹⁶ Ibid .Annexe 4

La première constatation concerne l'aspect de la distribution. Celle-ci est typiquement bimodale : nous retrouvons un premier mode sur l'intervalle centré sur la note de 20 sur 40 et un second centré celui de 28 sur 40. Cette répartition est typique d'un échantillon représentatif d'un mélange de deux populations différentes. Le premier groupe correspond aux postulants dont l'évaluation n'est pas convaincante et dont plus de la moitié sont ajournés. Le second regroupe clairement une population d'individus considérés comme disposant des qualités nécessaires à la formation d'IADE. Cette distribution présente le profil type d'une épreuve qui permet de sélectionner efficacement les candidats évalués sur les critères du jury.

La seconde réflexion concerne la première population dont le mode est centré sur la note de 20 sur 40. Cela peut être considéré comme une hésitation du jury à ajourner un candidat. Un axe d'amélioration, dans le sens d'un positionnement plus affirmé serait souhaitable. Celui-ci pourrait se concrétiser lors d'une séquence d'harmonisation des notes, en favorisant la comparaison à posteriori des candidats. Cela aurait pour effet de gommer le mode en redistribuant les notes de part et d'autre de la moyenne.

En résumé, il reste que l'épreuve possède une compétence de sélection tout à fait affirmée. La raison peut être attribuée à la composition d'un jury composé de professionnels confirmés, à même de saisir les valeurs professionnelles recherchées chez un postulant à la profession d'IADE.

Qu'en est-il maintenant du poids de cette évaluation dans la décision finale ?

Poids de l'oral dans la note finale

Cette fois-ci, nous allons rechercher le degré de corrélation qui pourrait exister entre la note attribuée à l'épreuve d'admission (l'entretien oral d'admission) et la note finale permettant le classement par ordre de mérite à l'issue de l'examen¹¹⁷. La figure ci-dessous présente pour chaque candidat, représenté par un point, la note attribuée à l'oral en abscisse (notes oraux) et en ordonnée celle obtenue au final (Notes finales).

¹¹⁷ Ibid Annexe 4

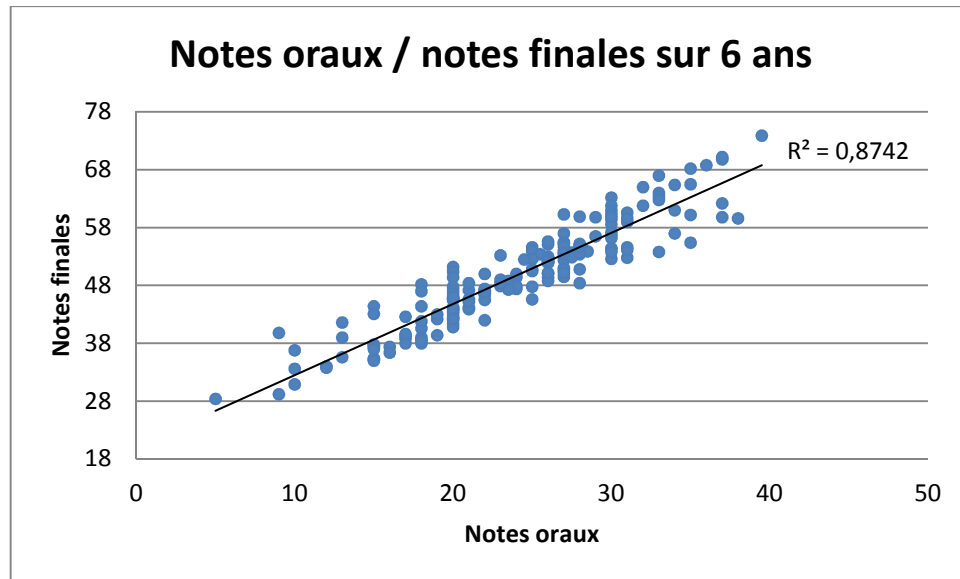


Figure 19: corrélation des notes d'oraux par rapport aux notes finales

Nous remarquons tout d'abord que le nuage de points est concentré autour de la droite de régression. Celle-ci est nettement ascendante sur le graphique. Ces deux paramètres d'appréciation visuelle sont confirmés par un coefficient R^2 de PEARSON égal à 0,87. Cette valeur particulièrement élevée permet de confirmer une corrélation hautement probable entre les deux notes.

En résumé, l'épreuve orale présente une qualité de sélection confirmée par la distribution de ses valeurs. Même si son pouvoir discriminant peut être amélioré en déplaçant le premier groupe en deçà de la moyenne, le rôle sélectif qui en est attendu est confirmé. De plus, la très nette corrélation de cette évaluation avec la note finale lui donne un poids majeur dans la décision finale de classement des candidats. Ces éléments confèrent à cette épreuve des qualités en totale opposition avec l'épreuve écrite et peuvent en faire un modèle de sélection efficace.

Il se pose maintenant la question d'un lien entre les deux évaluations. En d'autres termes peut-on penser qu'un candidat brillant à l'épreuve d'admissibilité écrite se verra attribuer une note en rapport lors de l'épreuve orale ? C'est ce que nous allons rechercher dans le chapitre suivant.

4.3.4. Il n'existe aucun lien entre l'épreuve écrite et l'épreuve orale

La corrélation entre l'épreuve écrite et l'épreuve orale est exprimée sur la figure suivante: Pour chaque candidat, représenté par un point, la note attribuée à l'écrit est représentée en abscisse (notes écrits) et en ordonnée celle obtenue à l'oral (Notes oraux).

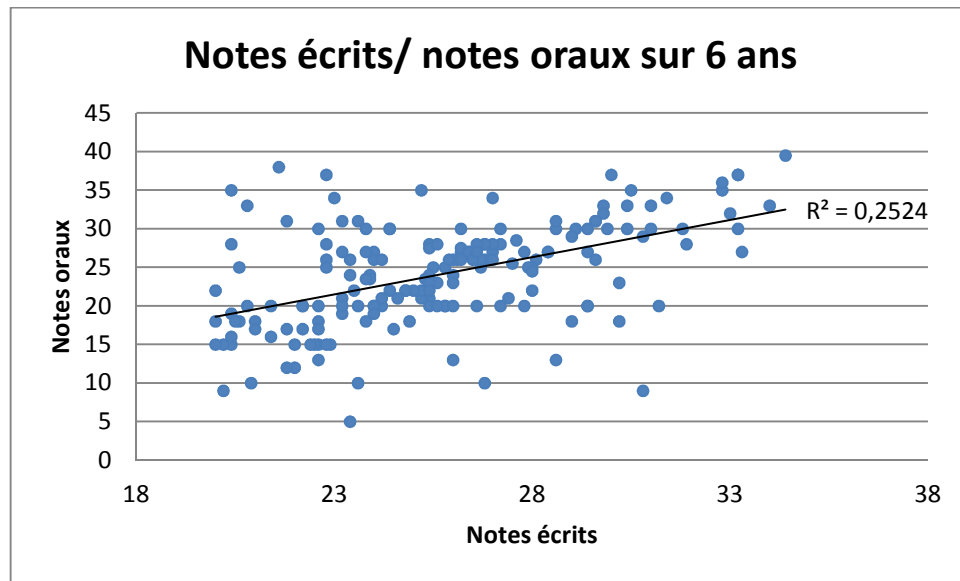


Figure 20: corrélation entre l'épreuve écrite et l'épreuve orale

Nous pouvons constater que la corrélation entre les deux épreuves n'existe pas. En effet, un coefficient R^2 égal à 0,25 n'est aucunement représentatif d'un lien entre les deux évaluations.

L'analyse que nous pouvons faire est assez simple. Les éléments de compétences et de connaissances recherchés sont différents dans les deux épreuves. Lors de l'admissibilité, seules les connaissances sont testées, en revanche, l'entretien d'admission recherche les compétences professionnelles et les aptitudes psycho-cognitives à suivre la formation.

En résumé, la discordance entre les résultats des deux épreuves n'a rien d'étonnant, en revanche, la répartition des notes finales observées peut être expliquée par cet effet de parasitage entre deux notes indépendantes dont la somme est le résultat de l'évaluation finale.

Ces éléments de réflexions étant énoncés, nous nous proposons d'en faire une synthèse apte à formuler une problématique étayée.

5. FORMULATION DE LA PROBLEMATIQUE

Avant de présenter les éléments permettant la rédaction de la problématique, nous rappelons la question de départ : "*Comment peut-on améliorer la sélection des candidats à suivre une formation lors d'un examen classant ou d'un concours ?*"

A la lumière des faits exposés et des résultats des différentes études présentées dans cette seconde partie de notre travail nous pouvons retenir cinq points essentiels :

- Les résultats obtenus en fin de formation confirment en majorité le classement au concours d'entrée. Ceci nous permet de penser que l'évaluation initiale des aptitudes à suivre la formation remplit son rôle et que la qualité de cette évaluation n'est pas à remettre en question. Le résultat final est conforme à nos attentes. D'aucun pourraient y voir un biais de Rosenthal au cours de la formation. Si toutefois c'est le cas, ce n'est pas un problème au regard du résultat obtenu.
- La répartition des notes finales, qui servent de support au classement des candidats, est satisfaisante du fait de sa distribution nettement décalée vers le haut de l'échelle de notation. Nous pouvons néanmoins lui reprocher un nombre un peu trop important de candidats ayant obtenu une note proche de la moyenne. Cela n'a que peu d'incidence, puisqu'il concerne les derniers classés sur la liste complémentaire. En effet ceux-ci ne sont que très rarement appelés à intégrer la formation. Nous pouvons peut-être y voir l'influence des résultats très peu sélectifs de l'épreuve écrite.
- Le résultat de l'entretien d'admission oral pèse de façon déterminante sur la décision finale. Ce n'est pas le cas de l'épreuve d'admissibilité, ce qui est confirmé par l'absence de lien entre les deux évaluations.
- La répartition des notes d'oral est tout à fait représentative d'une épreuve de sélection. Elle présente un pouvoir discriminant, au sens non péjoratif du mot, suffisant pour identifier la présence de deux groupes distincts, à savoir celui des candidats dont l'évaluation est jugée très satisfaisante et celui dont les prestations ont beaucoup moins convaincu les membres du jury. Remarquons à cet égard que le constat est surprenant, nous sommes en droit de penser qu'un examen oral est entaché de nombreux biais d'évaluation qui nuisent à son bon fonctionnement. Malgré cela, le résultat est exemplaire, aucune correction ne semble devoir être apportée.
- L'épreuve d'admissibilité écrite ne fait pas preuve des mêmes qualités. Une proportion importante des notes se situe autour de la limite d'admissibilité et leur distribution présente un aspect clairement gaussien incompatible avec la fonction de sélection du concours. De plus, les notes attribuées ne présentent que peu de lien (la corrélation est insuffisante) avec les notes finales permettant le classement des candidats. Le manque de qualité discriminante est particulièrement dommageable pour une épreuve d'admissibilité à l'entretien d'admission qui est prépondérant dans le classement final.

En conclusion, nous pouvons dire que l'épreuve écrite, dont le pouvoir discriminant est quasiment inexistant, mérite toute notre attention pour s'assurer qu'un candidat soumis à l'épreuve orale possède toutes les connaissances et les aptitudes fondamentales à suivre la formation.

En tenant compte de ces éléments résumant les observations préalablement développées, nous formulons la problématique de la façon suivante :

"Quelles améliorations peut-on apporter à la conception des épreuves écrites du concours d'entrée à l'école d'IADE afin d'en optimiser le pouvoir de sélection des candidats ? »

Pour répondre à cette question, il faut garder en mémoire qu'une épreuve n'est sélective que par les qualités discriminantes des éléments qui la compose. Dans ce sens, nous nous proposons, lors de la troisième partie de notre travail, de présenter, l'étude docimologique que nous avons réalisée portant sur le pouvoir discriminant des items de l'épreuve écrite et de leurs caractéristiques spécifiques.

Partie 3

III. La recherche : une étude docimologique portant sur le pouvoir discriminant des items de l'épreuve écrite

Ce travail de recherche se situe dans le champ de la docimologie éducatrice, il suit une démarche d'étude observationnelle visant à l'élaboration d'hypothèses. La vérification de celles-ci par l'expérimentation ne peut être réalisée dans le cadre temporel de ce mémoire. En effet, l'examen qui sert de support à notre étude se déroule une seule fois par an ; trois ou quatre années seraient donc nécessaires pour effectuer une série d'expérimentations exploitables. Ceci marque la limite d'une démarche qui se voulait initialement Bernardienne mais qui ne le sera que partiellement. Nous nous bornerons donc à une étude de type rétrospective qui pourra nous permettre de dégager des hypothèses susceptibles d'améliorer la qualité des épreuves écrites de sélection à l'entrée de l'école d'anesthésie.

1. LES BUTS DE L'ETUDE.

Les objectifs de cette étude sont de trois ordres : Tout d'abord rechercher parmi les questions proposées lors des sept dernières éditions de l'examen d'entrée à l'école d'IADE, lesquelles sont les plus discriminantes, c'est-à-dire disposant d'un indice de discrimination supérieur à 0,4. Le second est de définir les typologies auxquelles elles appartiennent et de repérer leurs caractéristiques spécifiques, si elles existent. Enfin, nous préciserons la proportion minimum de questions discriminantes à insérer dans l'épreuve écrite d'admissibilité.

2. L'ECHANTILLON EXPLOITE ET LES METHODES UTILISEES

En préambule, nous tenons à préciser que les données de l'année 2012 n'ont pas été prises en compte du fait de notre implication dans ce travail. Celles-ci auraient pu introduire un

biais méthodologique lié à l'attention que nous portons maintenant au choix des questions proposées.

2.1.Un échantillon représentant sept années de résultats à l'examen d'entrée

L'échantillon que nous avons étudié est composé des résultats obtenus aux épreuves écrites depuis l'année 2005 jusqu'à l'année 2011. La taille de l'échantillon (noté « N » par convention) est de 271 candidats. Chaque année, les postulants sont soumis réglementairement à vingt questions. Sur les sept années concernées, cela représente un nombre total de 5420 données. En regard du volume d'informations à traiter, nous avons choisi d'utiliser le tableur EXCEL[®] pour le tri et les calculs à effectuer. Cette étude est donc de type rétrospectif.

2.2.La méthodologie de l'étude

La méthodologie de cette étude repose sur quatre étapes successives :

- La première consiste à calculer les indices de discrimination, attribuant le caractère discriminant d'un item et celui de difficulté pour chaque question.
- La seconde est de repérer les items discriminants, ceux dont l'indice de discrimination est suffisamment élevé, et de les confronter à leur degré de difficulté afin d'établir s'il existe un lien entre les deux caractères.
- La troisième étape procède en un repérage des sessions annuelles qui présentent les plus faibles proportions de candidats dont les notes sont proches de la moyenne. Ce caractère de l'épreuve permet de repérer facilement celles qui laissent le moins de doute quant aux candidats à éliminer. Au sein de ces épreuves, les questions discriminantes seront extraites afin d'en établir la typologie et leurs éventuelles spécificités, ainsi que leur proportion sur les vingt items proposés chaque année.
- Enfin, après l'exposé des résultats, une discussion de ceux-ci visera à poser les hypothèses qui contribueront aux propositions de stratégie et d'ingénierie présentées dans la quatrième partie de notre travail.

Nous allons maintenant préciser les types d'indices utilisés, leurs interprétations et la justification de notre choix, avant de procéder à l'exposé des résultats.

2.2.1. Le choix de l'indice de discrimination

2.2.1.a. Les indices de discrimination

La littérature docimologique qui traite des problèmes éducatifs propose plusieurs indices de discrimination. Tous s'accordent sur le principe de mesure. Celui-ci consiste à faire la différence entre le pourcentage de réussite à une question du groupe de candidats dit « forts » et celui du groupe dit « faibles » à l'issue de la note de l'épreuve. La différence de méthode réside dans l'appréciation de ce qu'ils entendent sous les termes « forts » et « faibles » en terme de proportionnalité. Nous en avons retenu quatre parmi les plus représentatifs et les mieux documentés.

Nous commencerons par « L'indice D » de FINDLEY¹¹⁸. Celui-ci a été proposé par son auteur en 1956. C'est à notre connaissance le premier indice de discrimination défini. Celui-ci représente la différence entre le pourcentage de réussite du groupe fort (P^+) et ce même calcul pour le groupe faible (P^-) la formule est donc :

$$D = P^+ - P^-$$

Pour FINDLEY qui s'appuie sur les travaux de KELLEY (1939)¹¹⁹ : « *le groupe forts est constitué de ceux qui ont obtenu un score total qui se situe dans la catégorie des 27% supérieurs et le groupe faible dans la catégorie des 27% inférieurs* ». Cette valeur peut paraître surprenante mais elle est justifiée par la distribution Gaussienne couramment constatée, ce qui est le cas dans l'épreuve qui nous intéresse. L'indice D peut prendre une valeur comprise entre -1 et +1, la valeur 0 signifie que l'item a été réussi par autant de faibles que de forts. Si l'indice est négatif, cela signifie une plus grande réussite chez les faibles et permet d'émettre des doutes quant à l'opportunité d'une telle question. Plus tard EBEL (1965)¹²⁰ proposera une interprétation codifiée des valeurs de l'indice D :

Indice D = 0,40 ou plus	Item qui discrimine très bien
Indice D = 0,30 à 0,39	Item qui discrimine bien
Indice D = 0,20 à 0,29	Item qui discrimine peu
Indice D = 0,10 à 0,19	Item limite, à améliorer
Indice D = moins de 0,10	Item sans utilité réelle pour l'examen

Figure 21: interprétation des valeurs de l'indice D (EBEL-1965)

¹¹⁸ Findley W. G. (1956). « *A rationale for evaluation of item discrimination statistics* »

¹¹⁹ Kelley T. L. (1939). « *The selection of upper and lower groups for the validation of test items* », in *The Journal of Educational Psychology*, n°30, pp. 17-24.

¹²⁰ Ebel R.H. (1965) : « *Measuring educational achievement* »

Ce tableau nous permet de dire qu'une question présentant un indice D inférieur à 0,3 n'a aucun intérêt pour la construction d'une épreuve écrite de sélection. En effet, comme nous l'avons montré précédemment¹²¹, l'indicateur de qualité d'une épreuve de sélection est de mettre en évidence deux populations clairement identifiées de part et d'autre de la moyenne, et ceci n'est possible que par l'utilisation de questions discriminantes. Pour plus de sécurité, nous retiendrons la valeur de 0,4 dans le cadre de cette étude.

Nous allons maintenant présenter l'Index De Discrimination (IDD) proposé par GROOT (1972)¹²². Le principe de base est le même que pour FINDLEY mais l'expression mathématique est un peu différente :

$$\text{IDD} = \frac{\begin{array}{c} \text{(nombre de réponses correctes} \\ \text{dans le groupe de tête)} \end{array} - \begin{array}{c} \text{(nombre de réponses correctes dans} \\ \text{le groupe de queue)} \end{array}}{\mathbf{1/3 \text{ du nombre des étudiants}}}$$

Le calcul proposé par GROOT revient exactement au même, mais cette fois-ci, les groupes inférieurs et supérieurs représentent chacun un tiers des étudiants (33%) et non plus 27%, ce qui représente une différence considérable et modifie la valeur de l'indice. En effet cela élargi les groupes forts et faibles et minore par la même le contraste recherché. En effet, le premier groupe devient en moyenne moins fort et en le second, en moyenne moins faible. De ce fait, L'IDD perd en sensibilité en travaillant sur des groupes moins bien différenciés que dans l'indice D de FINDLEY. Nous notons que le choix de ce découpage n'est pas argumenté dans les documents que nous avons consultés.

A. QUINTON¹²³ propose un indice de discrimination basé sur la même répartition que G. GROOT en modifiant l'expression mathématique :

$$\text{Indice de discrimination} = 2 \times \frac{F + f}{N} \times 100$$

Où : « *F = nombre de réponses justes dans le groupe des forts, f = nombre de réponses justes dans le groupe des faibles et N = nombre total des étudiants appartenant aux groupes des forts et des faibles.* »

Le résultat de cette formule est strictement identique à celui de l'IDD de GROOT.

Enfin l'Institut de Recherche et de Documentation Pédagogique suisse (IRDP)¹²⁴ reprend pour sa part la définition et l'expression de l'Indice D de FINDLEY. Ce groupe de travail,

¹²¹ Op . cit. P. 60

¹²² Groot G. (1972) : « *Le problème de l'évaluation dans l'enseignement des langues vivantes* »

¹²³ Quinton A. (2005) : « *Docimologie* »

très orienté vers l'édu-métrie, cite également la possibilité d'utiliser la notion statistique d'écart-type. Si l'emploi de cet indicateur est effectivement possible, il n'existe pas, à notre connaissance, de table permettant d'interpréter le résultat de façon consensuelle.

L'exposé de ces différents indices (ou index) nous amène à faire le choix de l'indice D de FINDLEY, assujetti de l'interprétation de EBEL, pour deux raisons essentielles à nos yeux, à savoir : l'argumentation précise du choix de l'intervalle des groupes considérés comme « forts » ou « faible » créant un contraste suffisant entre les deux échantillons et par la même donne au calcul de l'indice une sensibilité supérieure. Vient ensuite le fait que le concept initial proposé par l'auteur n'a pas trouvé de contradicteurs majeurs.

L'inconvénient de tous ces indices est de faire référence à une « question réussie » par les candidats. Il nous faut, pour les utiliser, définir cette notion.

2.2.1.b. Qu'est ce qu'une question réussie ?

Trois choix s'offrent à nous : soit l'obtention de la moyenne est jugée suffisante, soit le maximum est exigé, soit enfin un moyen terme doit être trouvé.

A notre sens, l'obtention de la moyenne ne correspond pas au critère de réussite. En effet le doute qui subsiste toujours pour les valeurs limites ne répond pas à l'exigence d'une épreuve de sélection. L'attribution de la note maximale implique la perfection, aussi bien de la part du candidat que de l'examineur. Il est vrai qu'une simple faute d'orthographe, une écriture difficilement lisible peut entraîner la perte d'une partie certes infime de la cotation, mais suffisante pour que l'indice ne la prenne pas en compte. La solution d'un moyen terme nous semble plus appropriée. Le choix strictement intermédiaire de 75% de la cotation maximum, soit 7,5 sur dix, ne présente guère d'intérêt dans le sens où il est extrêmement rare, dans l'échantillon que nous allons étudier, que le barème de notation des questions permette d'attribuer une telle note.

Dans un souci de sécurité et pour éviter une sous évaluation du résultat, nous avons opté pour une note correspondant à 80% de la note maximum prévue pour la question. Ceci correspond à des notes de huit sur dix, ou seize sur vingt, ce qui reste en accord avec notre représentation de réussite à un item et correspond à la réalité des possibilités d'attribution des valeurs chiffrées.

¹²⁴ <http://www.irdp.ch/edumetrie/index.htm>

2.2.2. Le choix de l'indice de difficulté

Il nous reste à définir l'indice de difficulté. Celui proposé par FINDLEY et repris ensuite par d'autres auteurs est défini comme le pourcentage de réussite dans le groupe. Nous remarquons que plus ce pourcentage sera grand, plus le nombre de candidats ayant réussi sera grand. Ce constat nous amène à évoquer le paradoxe suivant : plus l'indice de difficulté est élevé, plus la question est facile ! Cet indicateur s'apparente donc plutôt à un indice de facilité. Cela n'enlève rien à sa pertinence, mais n'en simplifie pas la logique de lecture.

De plus il nous faut pouvoir vérifier la corrélation avec l'indice de discrimination, or celui-ci varie logiquement dans le sens de l'augmentation lorsque la qualité discriminante s'accroît, c'est à dire qu'il varie dans le sens inverse de l'indice de difficulté proposé ci-dessus.

Les chercheurs de l'IRDP se sont penchés sur ce problème et proposent un indice P dont la valeur augmente avec la difficulté, ce qui est plus naturellement lisible et devient utilisable dans la comparaison que nous ferons avec l'indice D. En voici l'expression mathématique :

$$P = 1 - \frac{m}{S_{max}}$$

Dans cette formule, m correspond à la moyenne des notes attribuées à l'item et S_{max} à la note maximum possible pour l'item. Nous prendrons pour exemples les cas extrêmes de réussite et d'échec à une question. Dans le premier cas, tous les candidats ont obtenu la note maximum, le rapport m / S_{max} est donc égal à 1. L'indice P sera dans ce cas égal à 0, ce qui est logique pour une question ne présentant aucune difficulté. Dans le second cas, aucun candidat ne s'est vu attribué de point à la question. Le rapport m / S_{max} est donc égal à 0. Le résultat du calcul de l'indice P sera donc cette fois-ci égal à 1. Ces exemples nous permettent de dire que l'indice P varie de 0 à 1, plus la difficulté est élevée, plus la valeur de l'indice P est proche de 1.

En résumé, le choix de l'indice D de FINDLEY nous paraît le plus approprié au regard de sa définition et de l'interprétation qui peut en être faite. L'indice P de difficulté des chercheurs de l'IRDP est également retenu pour ses qualités de lisibilité et d'aptitude à une étude de corrélation avec l'indice D propre à rechercher un lien entre ces deux caractéristiques.

La recherche d'un lien entre les deux indices est effectuée par le coefficient de corrélation R^2 de PEARSON précédemment utilisé.

L'argumentation des choix méthodologiques des indices retenu ayant été exposé, nous allons maintenant procéder à l'exposé des résultats de l'étude.

3. RESULTATS DE L'ETUDE

Dans ce paragraphe, nous allons exposer les différents résultats de l'étude et les commentaires qui leurs sont associés.

3.1. Pour faciliter la lecture : un exemple de la méthode concernant le repérage des indices de discrimination et de difficulté

L'indice D de discrimination et celui de difficulté ont été calculés pour chacune des questions soumises aux candidats¹²⁵. Ceux-ci ont préalablement été classés par ordre de mérite décroissant. Le groupe des forts est calculé sur la base de 27% des premiers (ramené au nombre entier le plus proche), le groupe des faibles représente 27% des derniers. **Nous donnerons pour exemple les résultats obtenus pour l'édition 2010 :**

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Indice de discrimination D (Findley)	0,82	0,18	0,45	0,73	0,64	0,36	0,18	0,18	0,09	0,45
indice de difficulté P (Edumétrie)	0,25	0,49	0,44	0,16	0,25	0,41	0,36	0,45	0,78	0,33
	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
Indice de discrimination D (Findley)	0,55	0,36	0,09	0,45	0,18	0,82	0,73	0,00	0,73	0,64
indice de difficulté P (Edumétrie)	0,61	0,54	0,31	0,32	0,07	0,52	0,52	0,96	0,50	0,24

Figure 22: exemple de résultats discrimination et difficulté

Pour plus de lisibilité des résultats exprimés dans le tableau précédent, nous fournirons un graphique de type polaire (voir figure suivante), où nous trouvons en périphérie le numéro de la question et sur son axe, représenté par le segment de droite qui le prolonge, les valeurs calculées de ses indices P et D.

¹²⁵ VOIR ANNEXE 5

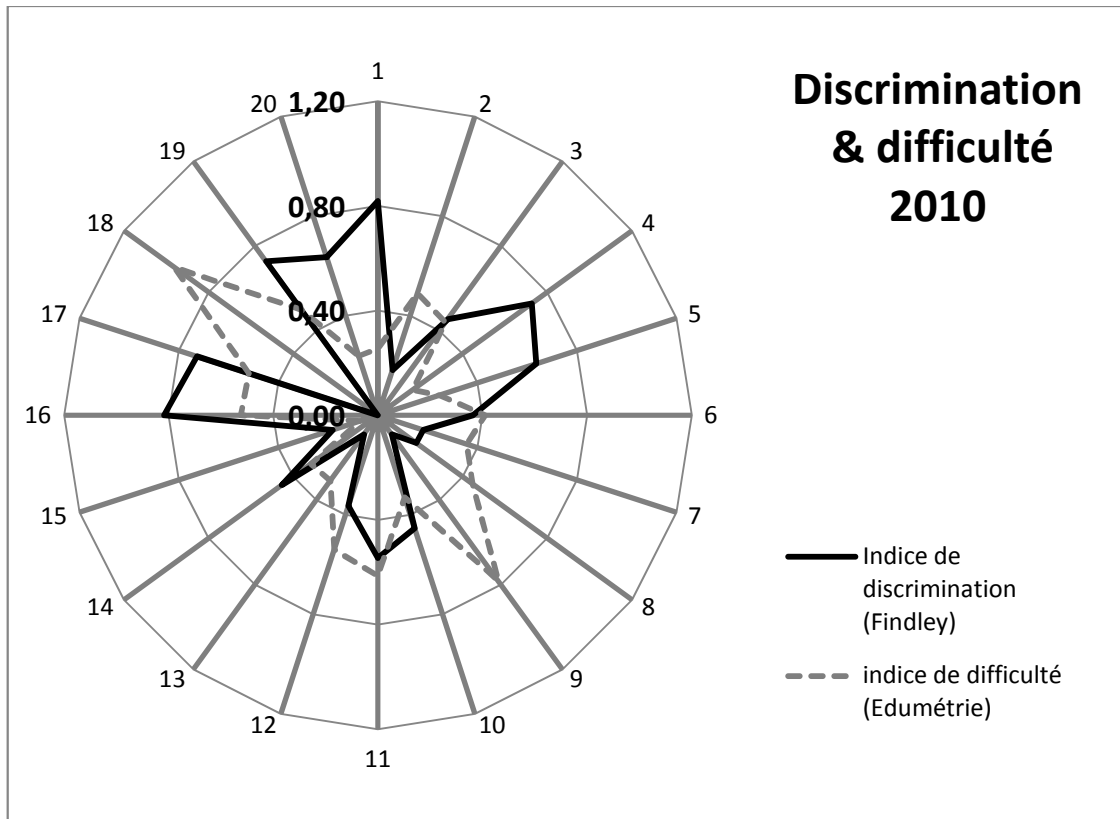


Figure 23: Exemple de diagramme polaire discrimination et difficulté

Cette dernière représentation permet une lecture rapide des résultats. Les questions les plus discriminantes apparaissent en « pointes » sur le tracé en trait plein noir au-delà du cercle représentant la valeur limite de l'indice D considérée de 0,40. Nous visualisons également la discordance entre les indices P et D pour une même question

3.1.1. Une question difficile n'est pas une question discriminante apte à la sélection.

Nous pouvons d'ores et déjà examiner la corrélation entre ces deux indices. Le graphique suivant en présente le résultat : Chaque point correspond à une question qui se voit attribuée son indice de discrimination en abscisse (axe « Indice D ») et son indice P de difficulté en ordonnée (axe « Indice de difficulté »). Pour exemple, le point entouré d'un cercle représente une question dont l'indice P est de 0,90 et l'indice D est de 0,13¹²⁶.

¹²⁶ Question n°7 session 2006 (Annexe 5)

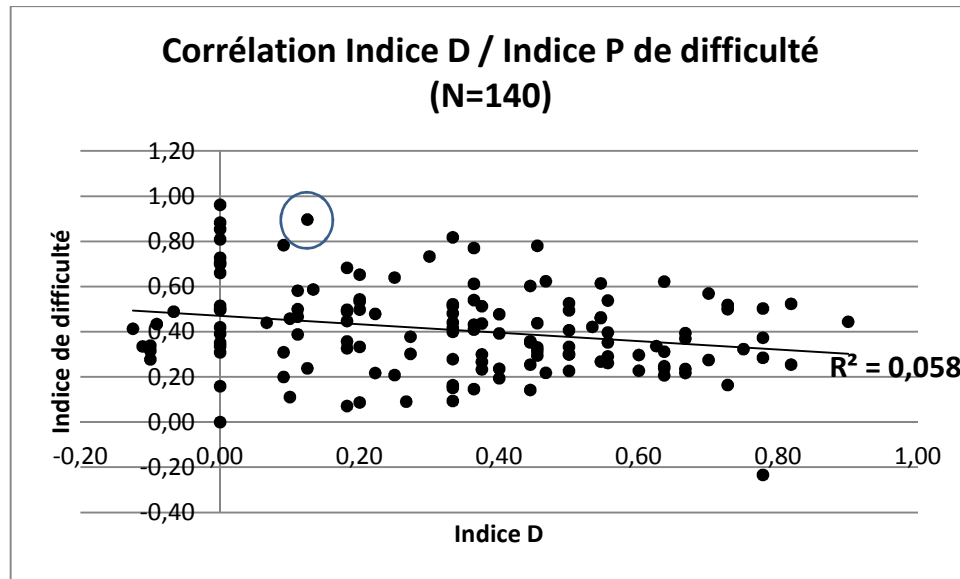


Figure 24: Corrélation indice D / indice de difficulté

Le test effectué sur les cent quarante questions de la série montre une disposition du nuage de points sans tendance particulière, illustrée par la droite de régression tracée sur le graphique. Ce constat visuel est confirmé par le calcul du coefficient de PEARSON dont la valeur de 0,058 signe une absence totale de lien.

Par ailleurs, **s'il est maintenant démontré qu'un item disposant d'un indice de difficulté important ne possède pas nécessairement de qualité discriminante, le degré de difficulté d'un item permet en revanche de réduire le nombre de candidats admissibles. Ceci fait tendre les résultats vers une courbe de Gauss dont le mode n'est pas déplacé mais qui présente une valeur plus faible. L'usage des questions affectées d'un indice P important est donc à réserver, pour adapter le niveau de difficulté de l'épreuve, quand le nombre de candidats inscrits est important.**

3.1.2. Plus le nombre des questions discriminantes est important, plus l'épreuve est sélective.

Le tableau suivant présente le nombre de questions chaque année pour lesquelles l'indice D atteint une valeur suffisante pour qualifier l'item de discriminant. Selon la codification de EBEL, nous avons étudié tous les items qui discriminent « au moins » biens les candidats, c'est-à-dire ceux dont la valeur de l'indice de discrimination est au moins égal à 0,3. Pour plus de précision, nous différencions ceux supérieurs à 0,3, 0,4 et 0,5 afin de

rechercher si leurs valeurs respectives sont prédictives d'une incidence sur le résultat escompté.

Année	Nombre de questions avec Indice D >0,5	Nombre de questions avec Indice D >0,4	Nombre de questions avec Indice D >0,3
2005	6	8	12
2006	8	8	13
2007	7	11	12
2008	1	5	10
2009	6	8	9
2010	8	11	13
2011	5	8	13

Figure 25: Nombre de questions discriminantes par année

Nous constatons des disparités importantes en fonction du niveau de l'indice D selon les années.

Si le nombre de questions dont l'indice de discrimination supérieur à 0,3 est à peu près homogène sur les sept années considérées, en revanche, à partir du seuil de 0,4, celui-ci devient particulièrement hétérogène au fur et à mesure que son niveau augmente. Lorsqu'il atteint le chiffre de 0,5, le constat est le même.

La question que nous devons nous poser maintenant est : Existe-t-il un lien entre le nombre de questions discriminantes et le pourcentage des notes situées dans un intervalle proche de la moyenne compris entre 9,5 et 10 ? En effet, si celui-ci est petit, 5, il y a de fortes chances que le nombre d'items discriminants ait une incidence sur cet intervalle. Pour cela, nous allons étudier le lien entre la valeur de leur indice D et le et le pourcentage de question ayant obtenues des notes entre 9,5 et 10,5. Le résultat de cette première approche est le suivant :

Année	% des notes entre 9,5 et 10,5/20:	Nombre de questions avec Indice D >0,5	Nombre de questions avec Indice D >0,4	Nombre de questions avec Indice D >0,3
2005	7,32	6	8	12
2006	13,33	8	8	13
2007	2,94	7	11	12
2008	21,82	1	5	10
2009	11,11	6	8	9
2010	4,76	8	11	13
2011	9,09	5	8	13

Figure 26: Valeur de D / intervalle autour de la moyenne

Nous remarquons que les années 2007 et 2010 ont un pourcentage des notes attribuées dans l'intervalle proche de la moyenne est particulièrement faible (2007 : 2,94% ; 2010 : 4,76%). Ceci nous permet de dire que la courbe de distribution des notes évoquées précédemment doit présenter le type bimodal recherché pour une épreuve de sélection. Dans chacun de ces cas, le nombre de questions dont l'indice D est supérieur à 0,4 est le plus important. Pour confirmer ou infirmer cette hypothèse, nous avons testé ce lien par le coefficient de SPEARMAN¹²⁷ pour les items présentant des indices supérieurs à 0,3, 0,4 et 0,5. Le tableau suivant présente les résultats :

Column A vs Column B	% des notes entre 9,5 et 10,5/20 vs Q discri >0,5	% des notes entre 9,5 et 10,5/20 vs Q discri >0,4	% des notes entre 9,5 et 10,5/20 vs Q discri >0,3
rs (Spearman)	-0,4364	-0,8964	-0,2620
P Value (one tailed)	0,1768	0,0062	0,2780
Was the pairing significantly effective?	No	Yes	No

Figure 27: test de SPEARMAN : indice D Vs % notes entre 9,5 et 10,5

L'hypothèse d'une corrélation significative entre le nombre de questions présentant un indice D supérieur à 0,4 et la proportion des notes situées dans l'intervalle compris entre 9,5 et 10,5 sur vingt est confirmée par une valeur de p (la probabilité que le lien soit dû au hasard) très nettement inférieure à 0,05 (voir la colonne grisée).

Par ailleurs et en relation avec les résultats exprimés sur la figure 27, nous ferons remarquer que la proportion de questions présentant un indice D supérieur à 0,5 devrait se situer aux environs de 75%, et sans dépasser cette limite. Ceci est confirmé par le test de SPEARMAN qui considère que leur présence exclusive dans l'épreuve écrite ne permet pas de dire qu'elle est suffisante pour limiter le nombre de notes autour de la moyenne.

Ce résultat nous permet également de penser qu'un nombre de onze questions, d'un niveau de discrimination de 0,4, sur les vingt proposées, est le minimum requis pour constituer une épreuve écrite de sélection convenable.

¹²⁷ Calcul réalisé avec le logiciel de statistiques PRISM®.

Nous allons maintenant nous intéresser plus spécifiquement aux millésimes 2007 et 2010 mis en évidence dans ce chapitre, en particulier aux vingt deux questions dont l'indice D est supérieur à 0,4. Ce travail est l'objet de ce que nous allons étudier dans le chapitre suivant.

3.2.Extraction des vingt deux questions discriminantes mises en évidence¹²⁸.

Parmi les résultats obtenus lors de l'étude portant sur les sept dernières années du concours, nous en avons particulièrement mis deux en évidence et leurs études méritent d'être approfondies. Il s'agit des sessions 2007 et 2010 dont l'examen des questions discriminantes servira de base, en partie, à nos préconisations en termes d'ingénierie de formation dans la suite de notre analyse. Comme convenu précédemment, les présentations des résultats chiffrés, sous forme de tableaux des millésimes 2007 et 2010, seront suivies de leurs représentations graphiques permettant une lecture plus visuelle ; elles seront traitées de la même manière.

Voici dans un premier temps les résultats de l'année 2007 dont nous rappelons qu'elle a été retenue pour son faible pourcentage (2,94 %) de candidats ayant obtenu des notes entre 9,5 et 10,5 sur vingt. Ceci répond aux critères d'une épreuve discriminante, c'est-à-dire la capacité de choisir clairement les postulants à la formation en évitant une évaluation « *ni oui - ni non* » exprimée par une note très proche de la moyenne et révélatrice d'une indécision des examinateurs :

2007	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Q 17	Q 18	Q 19	Q 20
Indice D de discrimination (Findley)	0,78	0,56	0,44	0,11	0,33	0,11	0,56	0,22	0,44	0,44	0,00	0,22	0,67	0,67	0,11	0,00	0,56	0,78	0,00	0,56
Indice P de difficulté (Edumétrie)	0,37	0,35	0,60	0,39	0,44	0,50	0,26	0,22	0,35	0,36	0,70	0,48	0,24	0,22	0,34	0,71	0,54	0,29	0,31	0,29

Figure 28: Indices D et P des questions 2007

¹²⁸ Voir ANNEXE 6

Nous constatons qu'il existe onze questions dont l'indice D est compris entre 0,44 et 0,78. Ces chiffres correspondent à des « *items qui discriminent très bien* » selon l'échelle d'EBEL permettant l'interprétation des résultats du calcul de l'indice de discrimination. Les questions concernées sont mises en évidence par une nuance grisée plus intense. Nous remarquerons, une fois de plus, que les indices P de difficulté ne sont pas en accord avec le caractère discriminant des questions proposées, ce qui confirme les résultats précédemment observés lors de l'étude générale malgré le choix ciblé d'items à fort pouvoir discriminant. La figure suivante visualise les composants de l'épreuve qui lui confèrent sa compétence de sélection.

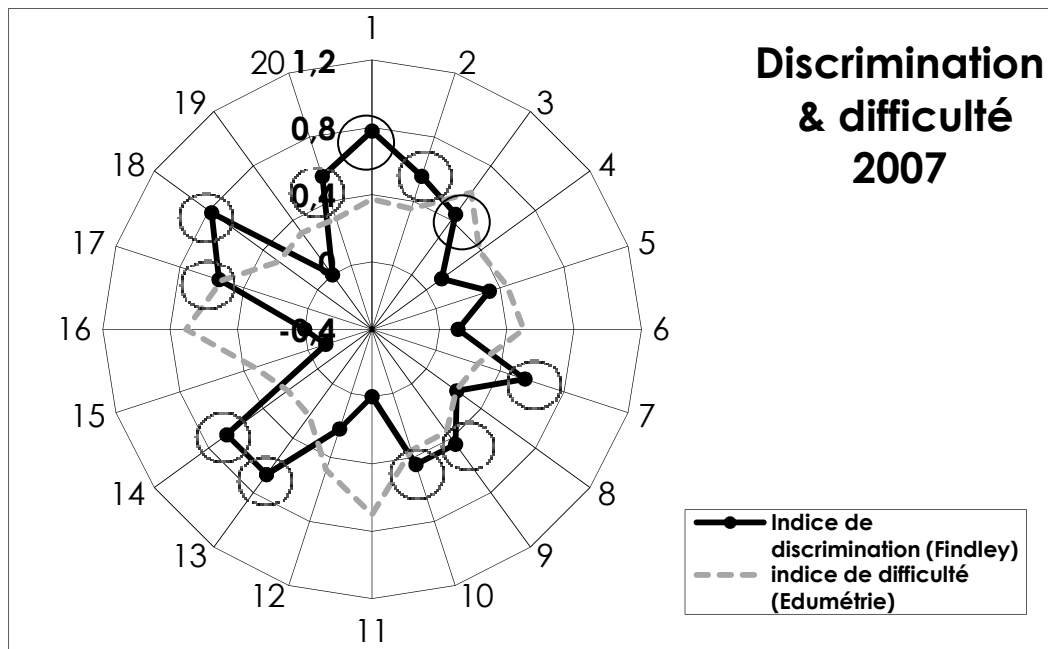


Figure 29: Représentation graphique des Indices D et P des 20 questions 2007

Nous avons identifié les questions dont l'indice D est supérieur à 0,4 en les entourant d'un cercle. Il s'agit des items :

1, 2, 3, 7, 9, 10, 13, 14, 17, 18 et 20.

Pour des commodités de repérage dans les annexes, nous les nommerons par un premier chiffre représentant l'année de la session de l'examen (7 pour l'année 2007 et 10 pour l'année 2010), suivi d'un trait d'union et du numéro de la question dans la dite session, cela nous conduit à la nomenclature suivante:

7-1, 7-2, 7-3, 7-7, 7-9, 7-10, 7-13, 7-14, 7-17, 7-18 et 7-20

Nous constatons qu'il existe 11 questions affectées d'un indice D de discrimination supérieur à 0,4 dont 8 se voient attribuer un score supérieur à 0,5

Un exemple de chaque type de question selon la classification retenue sera fourni lors de la discussion à la fin de l'exposé des résultats de ces deux années.

Dans un second temps, nous présentons les résultats de l'année 2010 dont nous rappelons qu'elle a été retenue pour son faible pourcentage (4,76 %) de candidats ayant obtenu des notes entre 9,5 et 10,5 sur vingt. Ceci répond aux critères d'une épreuve discriminante, c'est-à-dire la capacité de choisir clairement les postulants à la formation en évitant une évaluation « *ni oui - ni non* » exprimée par une note très proche de la moyenne et révélatrice d'une indécision des examinateurs :

2010	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Q 17	Q 18	Q 19	Q 20
Indice D de discrimination (Findley)	0,82	0,18	0,45	0,73	0,64	0,36	0,18	0,18	0,09	0,45	0,55	0,36	0,09	0,45	0,18	0,82	0,73	0,00	0,73	0,64
indice P de difficulté (Edumétrie)	0,25	0,49	0,44	0,16	0,25	0,41	0,36	0,45	0,78	0,33	0,61	0,54	0,31	0,32	0,07	0,52	0,52	0,96	0,50	0,24

Figure 30: Indices D et P des questions 2010

Nous constatons qu'il existe onze questions dont l'indice D est compris entre 0,45 et 0,82. Ces chiffres correspondent à des « *items qui discriminent très bien* » selon l'échelle d'EBEL permettant l'interprétation des résultats du calcul de l'indice de discrimination. Les questions concernées sont mises en évidence par une nuance grisée plus intense. Nous remarquerons, une fois de plus, que les indices P de difficulté ne sont pas en accord avec le caractère discriminant des questions proposées, ce qui confirme les résultats précédemment observés lors de l'étude générale malgré le choix ciblé d'items à fort pouvoir discriminant.

La figure suivante visualise les composants de l'épreuve qui lui confèrent sa compétence de sélection.

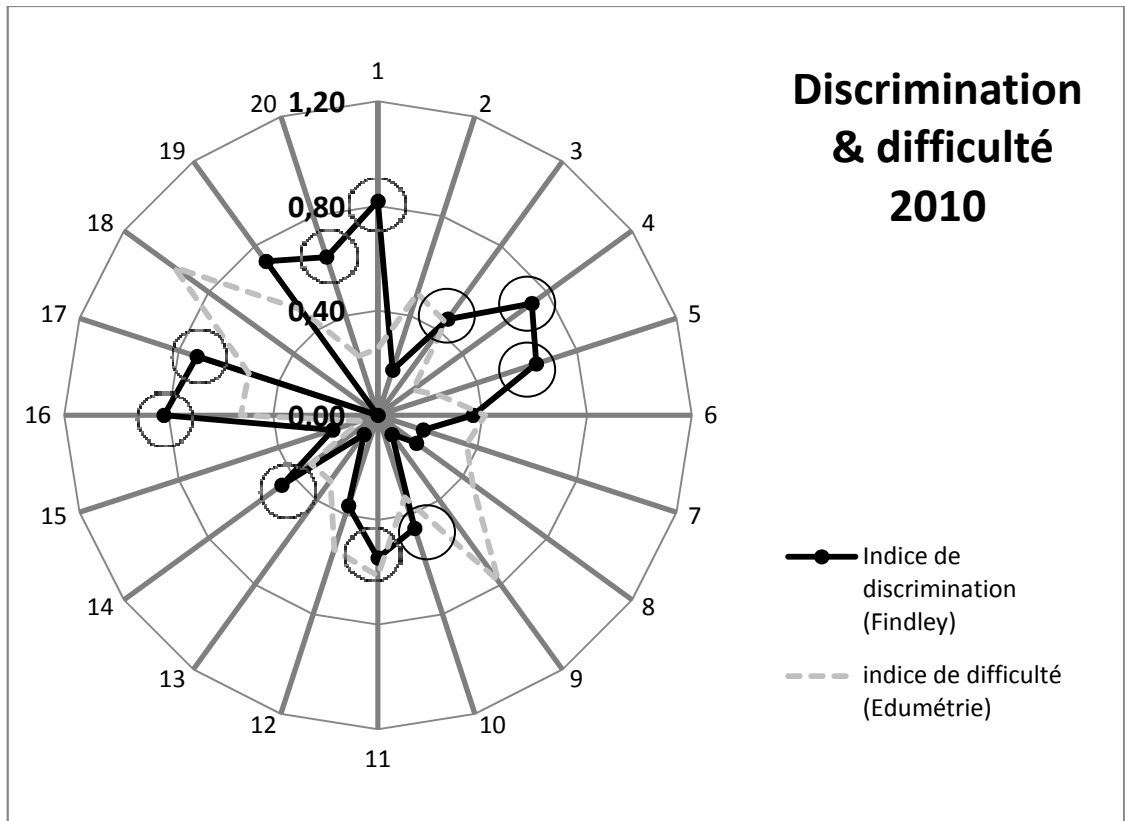


Figure 31: Représentation graphique des Indices D et P des 20 questions 2010

Nous avons identifié les questions dont l'indice D est supérieur à 0,4 en les entourant d'un cercle. Il s'agit des items :

1, 3, 4, 5, 10, 11, 14, 16, 17, 19 et 20

Pour des commodités de repérage dans les annexes, nous les nommerons une nouvelle fois par un premier chiffre représentant l'année de la session de l'examen (7 pour l'année 2007 et 10 pour l'année 2010), suivit d'un trait d'union et du numéro de la question dans la dite session, cela nous conduit à la nomenclature suivante:

10-1, 10-3, 10-4, 10-5, 10-10, 10-11, 10-14, 10-16, 10-17, 10-19 et 10-20

Nous constatons une seconde fois qu'il existe 11 questions affectées d'un indice D de discrimination supérieur à 0,4 dont également 8 se voient attribuer un score supérieur à 0,5

Un exemple de chaque type de question selon la classification retenue sera fourni lors de la discussion, à la fin de l'exposé des résultats de ces deux années.

Nous avons maintenant repéré les items les plus représentatifs des épreuves de sélection, il nous faut à présent les catégoriser pour pouvoir utiliser ces résultats en termes de préconisation d'usage.

3.3. Une référence nécessaire de classification des questions

La catégorisation des items écrits a fait l'objet de nombreux travaux. La référence la plus actuelle est la synthèse réalisée par D. LECLERCQ¹²⁹ en 2005. Celle-ci fait état de cinq types de question destinées à évaluer le niveau de connaissance des étudiants. Dans le cas qui nous intéresse, il s'agit de professionnels candidats à l'entrée en formation spécialisée mais cela ne change rien quant à l'objectif des questions et à la catégorie à laquelle elles appartiennent. Dans ce champ, nous distinguerons :

- **Les Questionnaires à Choix Multiples couramment appelés QCM.** Il s'agit de poser une question et de fournir au candidat plusieurs choix de réponses, le plus souvent trois ou quatre. Celui-ci doit cocher celle (ou celles) qui lui paraît (paraissent) être la (les) plus adaptées. Dans cet exercice, il est en général mentionné que la bonne réponse est unique, multiple ou éventuellement non mentionnée. La possibilité que ces trois variations soient présentes lors d'un même questionnement est très rare. Depuis bien longtemps, les QCM suscitent des commentaires défavorables par les défenseurs de la question ouverte, mais il faut bien reconnaître que les biais d'évaluation sont beaucoup mieux maîtrisés par ce type d'item. Viennent ensuite les critiques concernant l'adaptation des étudiants à ce genre d'évaluation. Nous verront qu'il existe des parades méthodologiques à la pratique de ce que les anglo-saxons nomment le « guessing »¹³⁰ dont nous reparlerons plus tard.

- **Les questionnaires « VRAI-FAUX »,** qui sont des cas limites de QCM et souvent à juste titre intégrés dans cette catégorie. En effet il s'agit dans ce cas de choisir entre deux possibilités de réponses, c'est donc bien un choix multiple, mais le fait que la possibilité soit binaire implique une « solution adaptative » de la part des candidats ne connaissant pas le sujet, du type tirage à pile ou face.

- **Les Questionnaires à Choix Large également ou QCL,** ceux-ci sont construits de la même manière que les QCM, mais le nombre de possibilités de réponses peut varier d'une dizaine à plusieurs centaines. Dans ce dernier cas le panel de réponses possible peut être unique pour l'ensemble des questions posées. C'est une méthode efficace pour limiter le

¹²⁹ Leclercq D. (2005) « *Edumétrie et docimologie pour praticiens chercheurs* »

¹³⁰ En français : « en devinant, en supposant »

« guessing » mais le temps nécessaire à la lecture de l'énoncé est très nettement augmenté et limite d'autant le nombre d'items à proposer dans le temps réglementaire.

- Les Questionnaires à Réponses Courtes ou de longueurs Moyennes dénommés CROC ou CROM : l'attendu est dans ces cas un mot, ou une phrase très courte. Dans cette catégorie nous retrouvons les schémas et les calculs dont nous attendons simplement les résultats, sans faire apparaître la méthode (seul le résultat compte). Dans le cas des CROM, la longueur de la réponse peut tendre vers une question à réponse ouverte.

- Les Questionnaires à Réponses Ouvertes (QRO)

Les questions ouvertes, de type rédactionnel, ne sont pas représentées dans cette étude, en effet, celles-ci sont absentes des épreuves que nous avons étudiées. La nouvelle réglementation actuellement en préparation, demandera à priori une question de composition destinée à tester les compétences rédactionnelles des futurs étudiants. Celle-ci nous réclamera un complément d'approche docimologique car elle pose le problème des biais d'évaluations majorés par la rédaction et l'interprétation qui peut en être faite. Une étude spécifique complémentaire sera alors nécessaire pour apprécier la méthodologie d'évaluation que nous devons en faire.

Les définitions des catégories étant précisées, nous pouvons maintenant classer les items extraits des épreuves de 2007 et 2010 et en faire une analyse.

4. DISCUSSION :LES QCM ET LES CROC SONT LES PLUS PERFORMANTS.

Les vingt deux questions discriminantes extraites de nos données ont été classées en trois grandes catégories : dix sont des Questions à Choix Multiple, dix sont des Questions à Réponses Ouvertes Courtes et un seul item est une Question à Réponse Ouverte Moyenne.

Pour des raisons pratiques, et avec l'objectif d'affiner le tri, les QCM ont été divisés en trois sous catégories :

- les QCM généralistes (5 items),
- les questions inhérentes à la transfusion sanguine pour l'importance majeure qui leur est accordée par le jury (4 items),
- et enfin, les « vrai – faux » pour leur spécificité exprimée dans le chapitre précédent (1 item).

Nous avons procédé de la même manière pour les QROC en distinguant également trois sous catégories :

- les items généralistes (1 item),
- les questions de calculs pour l'importance professionnelle à leur accorder 3 items),
- et enfin les schémas d'anatomie en raison de leur grande fréquence d'apparition dans cette sélection (6.items)

Afin de mieux visualiser la typologie de ces items, nous proposons un exemple choisi de question discriminante représentative de chacune de ces catégories.

4.1.Un exemple de question discriminante par types retenus

QCM « généraliste » :

10-1 Cochez la case correspondant à la bonne réponse.

...../ 10 pts

- a- La formation de glucose à partir d'acide aminés est :
 - La glycogénolyse ☐
 - La néoglucogénèse ☒
 - La glycogénèse ☐
- b- Le volume de sang contenu dans les veines est de l'ordre de :
 - 40% du sang de l'organisme ☐
 - 75% du sang de l'organisme ☐
 - 60% du sang de l'organisme ☒
- c- L'air contient :
 - 51% d'oxygène ☐
 - 21% d'oxygène ☒
 - 25% d'oxygène ☐
- d- La stimulation du nerf vague provoque :
 - Une augmentation de la pression artérielle ☐
 - Une augmentation de la fréquence cardiaque ☐
 - Une diminution de la fréquence cardiaque ☒
- e- La concentration en pourcentage d'une solution contenant 0,08g/ml est de :
 - 0,08% ☐
 - 0,8% ☐
 - 8% ☒
- f- Dans le foie la rénine transforme :
 - L'angiotensine I en angiotensine II ☐
 - L'angiotensinogène en angiotensine I ☒

- L'angiotensine II en aldostérone ☐
- g- Le L.C.R est contenu au niveau de :
 - L'espace sous arachnoïdien ☒X
 - L'espace sous dural ☐
 - L'espace péri-dural ☐
- h- La C.R.F (capacité résiduelle fonctionnelle) est égale à :
 - volume résiduel + volume de réserve inspiratoire ☐
 - volume courant + volume de réserve expiratoire ☐
 - volume résiduel + volume de réserve expiratoire ☒X
- i- le nombre de paires racines nerveuses d'origine rachidienne est :
 - 12 ☐
 - 25 ☐
 - 31 ☒X
- j- Le test d'exploration de l'hémostase primaire est :
 - Le temps de saignement ☒X
 - Le temps de céphaline + activateur ☐
 - Le dosage de fibrinogène ☐

QCM « vrai – Faux »

7-2 Pour répondre aux 10 affirmations du tableau ci-dessous, mettez une croix dans les cases correspondant aux items VRAI ou FAUX.






















...../ 10 pts

AFFIRMATIONS	VRAI	FAUX
La calcémie n'est pas élevée dans l'insuffisance parathyroïdienne.	X	
Le temps de coagulation correspond au temps de formation du clou plaquettaire.		X
La kaliémie est abaissée au cours des états de chocs.		X
Au cours des diarrhées prolongées, la natrémie peut être élevée.	X	
La vitesse de sédimentation indique la vitesse de croissance des globules blancs.		X
La réserve alcaline est augmentée dans l'acidose métabolique.		X
Chez le nouveau-né, la crise hémolytique favorise l'augmentation du nombre des hématies.		X
Le neurotransmetteur des jonctions neuromusculaires des muscles striés est la succinylcholine.		X
Toute diminution de la PaO ₂ inférieure à 60 mmHg entraîne une diminution de la ventilation.		X
Chez l'adulte, la pression normale du liquide céphalo-rachidien est comprise entre 0 et 5 cm H ₂ O.	X	

QCM « transfusion »

10-16 Parmi les sept possibilités de contrôles pré-transfusionnels proposés ci-dessous, indiquez les groupes A, B, AB et O dans les cases de la colonne « GROUPES ».

...../ 10 pts

GROUPES	Sérum anti-A	Sérum anti-B	Sérum anti-AB
<i>AB</i>			
			
<i>A</i>			
<i>B</i>			
			
<i>O</i>			
			

QROC « généraliste »

7-7- Complétez le tableau suivant :

...../ 10 pts

Paramètre sanguin	Valeur normale chez l'homme
Globules rouges	4,2 à 4,7 million /mm ³
Globules blancs	4000 à 10000 / mm ³
Hématocrite	40 à 52 %
Hémoglobine	13 à 18 g /100ml
Plaquettes	150000 à 350000 / mm ³
% de prothrombine	80 à 100 %
Créatinine	7 à 13 mg/l ou 40 à 120 µmol/l
Sodium	138 à 142 mmol/ ou mEq/l
Potassium	3,5 à 4,9 mmol/l ou mEq/l
Bicarbonates	25 28 mmol/l ou mEq/l

QROC « calculs »

7-10 Vous êtes infirmier(ère) en salle de surveillance post-interventionnelle. Nous sommes le 1^{er} janvier, il est 13 heures. Le médecin prescrit une perfusion continue de NAROPENE® à 2^o/_{oo} sur le cathéter d'anesthésie loco-régionale d'un patient de 80 kg, à raison de 100µg/kg/h. Le médicament est conditionné dans une poche de 100 ml.

...../ 10 pts

a) Quel sera le débit de perfusion ?

4 ml/h

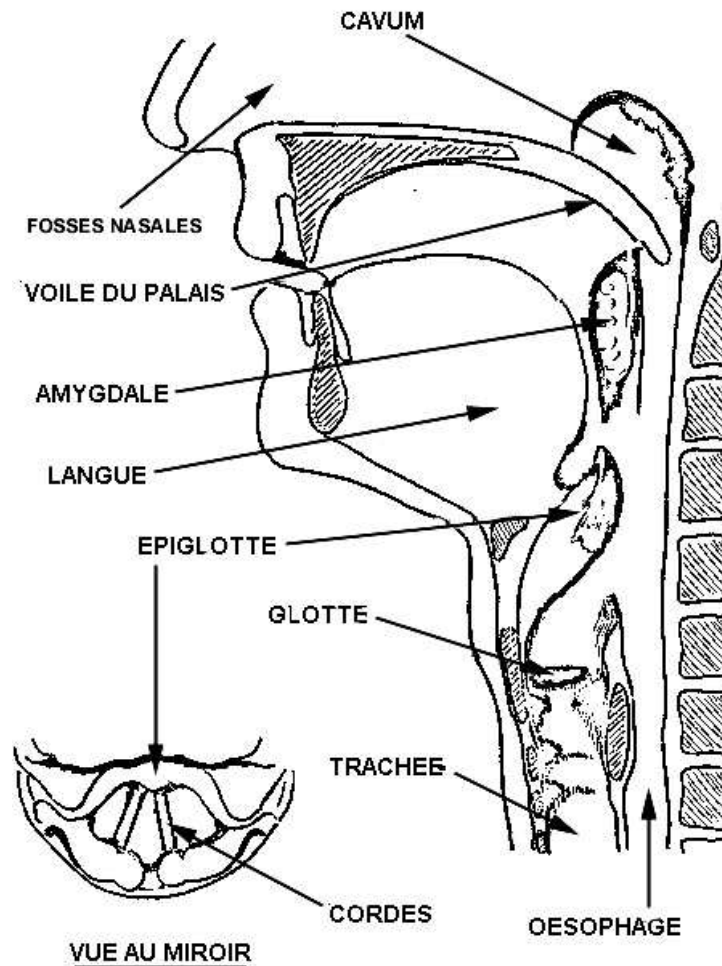
b) A quelle date et à quelle heure pouvez-vous prévoir la fin de cette perfusion ?

Le 2 janvier à 14 heures

QROC « schémas »

7-1 Légendez le schéma suivant en indiquant vos réponses dans les dix cadres prédéfinis ci-dessous.

...../ 10 pts



QROM

7-3 La loi n° 2002-303 du 4 mars 2002 « relative aux droits des malades et à la qualité du système de santé », donne aux patients un certain nombre de droits : citez-en deux.

...../ 10 pts

- *l'accès au dossier médical.*
- *les nouveaux droits des mineurs concernant leur secret médical vis à vis de l'autorité parentale.*
- *renforcement de la place des associations de patients et usagers dans la gestion du système de santé.*
- *suspension des peines pour les condamnés très malades.*
- *l'indemnisation des accidents médicaux.*

4.2. Analyse de la performance implicite des typologies des questions retenues

Le tableau suivant permet de faire la synthèse de l'étude des questions discriminantes misent en évidence par notre travail préliminaire. Nous présentons dans les deux premières colonnes les typologies générales et leurs sous catégories. Les deux colonnes suivantes expriment le nombre de questions non discriminantes du même type présentes dans les épreuves de 2007 et 2010; ainsi que le rapport entre le nombre de questions discriminantes correspondant à une catégorie et le nombre total d'items du même type non retenus en regard de leur pouvoir discriminant trop faible.

L'objectif de ce travail est de montrer que la typologie principale d'une question n'est pas le seul critère objectivable de son pouvoir discriminant et qu'il existe des sous critères pouvant expliquer son importance dans la sélection des candidats.

Pour cela nous avons calculé pour chaque sous type de questions présentes dans les épreuves étudiées, le pourcentage d'items discriminants par rapport à ceux qui ne le sont pas.

Le résultat de cette étude est le suivant :

Typologie des questions	Questions discriminantes		Nombre de questions <u>non discriminantes</u> du même type	Rapport questions discriminantes/ Total questions du même type	
QCM	Généralistes	7-9 / 10-1 /10-4/ 10-5 / 10-13	2	5/7 71,5 %	10/16 62,5 %
	Transfusion	7-13 / 7-14 / 10-16 / 10-20	0	4/4 100 %	
	Vrai faux	7-2	4	1/5 20 %	
QROC	généralistes	7-7	3	1/4 25 %	10/18 55,5 %
	Calculs	7-10 / 10-17 / 10-19	2	3/5 60 %	
	Schémas	7-1 / 7-17 / 7-18 / 7-20 / 10-10 / 10-14	3	6/9 66,6 %	
CROM	7-3 / 10-11		4	2/6 33,3 %	

Figure 32: Tableau récapitulatif des questions discriminantes

Nous constatons que la réalisation d'un QCM permet d'obtenir dans plus de soixante pour cent des cas à un item très discriminant. Les réalisations les plus probantes sont les questions concernant la transfusion (100 % d'efficacité), ce qui est très intéressant sur le plan professionnel, et les QCM généralistes qui représentent lors de leur expression, plus de soixante dix pour cent d'items discriminants.

Comme nous pouvions nous y attendre, les questions de type « vrai – faux » ne sont pas systématiquement représentatives d'une épreuve de sélection, en effet, seules vingt pour cent d'entre elles obtiennent un indice de discrimination supérieur à 0,4. A ce sujet, N. RICHE¹³¹ met l'accent sur la facile dérive de rédaction des questionnaires « oui – non » en précisant que leur généralisation d'emploi est :

¹³¹ Nicole RICHE in D. LECLERCQ : Op Sit.

"Une maladie qui existe à l'état endémique chez tout formateur ... Je me refuse à mettre au monde le sigle tiré de Question Oui Non qui viendrait agrandir la famille QCM, QROC, etc. par crainte que ce sigle plaise et contribue à la propagation de la maladie au lieu de l'éradication visée"

Si les QCM limitent les écarts de notation, ils ne suppriment pas totalement les risques d'erreurs humaines. Une solution possible est de proposer une correction instrumentée par un système informatique de lecture optique des réponses.

Par ailleurs, il ne faut pas oublier que les candidats s'adaptent rapidement à la typologie des Questions à Réponses Multiple. Nous pouvons constater l'usage d'une réponse adaptative que les anglo-saxons nomment « guessing », c'est-à-dire la capacité à répondre à un QCM en exploitant toutes les informations disponibles dans l'énoncé, sans forcément connaître le sujet, en pariant sur la probabilité d'une réponse exacte. Une possibilité également exploitée est celle du « blind guessing », à savoir la réponse au hasard qui peut être tout à fait efficace dans le cadre des questionnaires « vrai – faux ». En effet, sur dix questions posées en tableau, une réponse positive dans tous les cas est statistiquement prédictive d'une réussite à cinquante pour cent : c'est le problème du « pile ou face » !

En regard de ces possibilités adaptatives, la solution qualifiée de « notation multiplicatrice » par D LECLERCQ peut être envisagée. Elle consiste en une note tenant compte d'un minimum requis de réponses exactes, en deçà duquel aucune note supérieure à zéro ne sera attribuée.

Signalons enfin les possibilités de « Solution Générale Implicite » (SGI) qui permettent d'intégrer des réponses absurdes éliminatoires, ou bien la possibilité qu'aucune réponse ne soit envisageable. Dans ce dernier cas, il est possible de proposer la solution « aucune » ou « autre », mais ceci n'est pas obligatoire : l'option de non réponse peut être proposée au candidat de façon implicite, mais cela induit un biais de correction de la part de l'examineur qui ne pourra pas juger de la qualité d'une non réponse. C'est en partie le cas des questions de transfusion sanguines présentées ci-dessus qui font apparaître des solutions absurdes, mais si toute fois le candidat se trompe, le jugement sera négatif et se soldera par un zéro.

Les Questionnaires à Réponses Ouvertes Courtes sont également intéressants. Parmi ceux-ci, nous retiendrons surtout les items concernant les schémas anatomiques et les calculs qui sont à plus de soixante pour cent représentatifs de questions discriminantes. En revanche, les QROC généralistes sont potentiellement peu discriminants (25 %).

Les Questionnaires à Réponses Ouvertes Moyenne sont certes assez intéressants, mais leur fiabilité de conception reste sujette à précautions puisqu'ils ne sont discriminants que dans un tiers des cas. Ceci peut être mis en lien avec la possibilité de réponses plus longues que prévues de la part des candidats, ce qui engage une évaluation subjective du jury confronté à des réponses plus proches de celles attendues pour des questions ouvertes.

Initialement, nous avions l'intention d'étudier l'influence du pas de notation des items. La raison était de considérer qu'une note égale à la moyenne pouvait influencer sur le résultat final. Il est évident que cela ne peut en aucun cas influencer sur le résultat final dans le sens où d'une part, le simple fait qu'une question soit discriminante implique qu'elle sélectionne naturellement les bons des moins bons, et que d'autre part la possibilité qu'une note supérieure à la moyenne attribuée à un item peut être compensée une appréciation équivalente négative d'une autre question.

5. CRITIQUE ET LIMITES ET DE L'ETUDE

Ce travail a été effectué sur une seule école, ce qui ne permet pas de le généraliser à l'ensemble aux autres établissements du même type. Nous mettons en avant que chaque école propose des orientations et des projets pédagogiques spécifiques qui génèrent une diversité dans la formation des professionnels, sur les plans des valeurs et des éléments d'enseignements. Par ailleurs, il existe une forme de concurrence entre établissements, qui ne permet pas la coopération nécessaire à une extension multi centrique de ce type d'étude. Ceci est d'autant plus dommageable que cela pourrait accélérer la phase d'expérimentation qui doit y succéder.

D'autre part, une étude statistique du lien entre les notes proches de la moyenne et les questions dont l'indice de discrimination est significatif aurait pu faire l'objet d'un test de χ^2 corrigé par le V de Cramer au regard du faible nombre de données. Nous ne maîtrisons pas ce type de test et nous avons choisi de nous abstenir des conclusions erronées qui pourraient en découler.

6. CONCLUSION

En conclusion, nous pourrions considérer que dans le cadre de la conception d'items discriminants, les QCM sont très intéressants. Nous retiendrons en particulier ceux concernant la transfusion sanguine et les questions plus généralistes. En revanche les

« vrai – faux » sont à proposer avec beaucoup de précautions en regard des attitudes adaptatives des candidats qui se présentent aux concours.

Les Questionnaires à Réponses Ouvertes Courtes présentent un intérêt certain, en particulier pour les schémas d'anatomie et lorsqu'il s'agit de questions de calculs.

Les Questionnaires à Réponses Ouvertes de longueur Moyenne doivent en revanche faire l'objet de précautions particulières, de façon à éviter une dérive vers les questions ouvertes et l'évaluation subjective qui peut en être faite par le jury.

Nous pouvons maintenant faire état des préconisations, en termes de stratégie et d'ingénierie de conception des épreuves écrites, lors d'un examen classant permettant l'entrée en formation d'Infirmiers Anesthésistes

Partie 4

IV. Propositions d'améliorations et possibilités de réinvestissement professionnel

1. PROPOSITIONS D'AMELIORATION DE L'EPREUVE ECRITE

Les propositions d'amélioration sont faites d'un point de vue distancié, universitaire, à partir des éléments étudiés et montrés dans les parties deux et trois de ce mémoire et ne tiennent pas nécessairement compte des obligations réglementaires. Toutes ont pour objectif d'améliorer le pouvoir discriminant de l'épreuve écrite, afin de fiabiliser la sélection des candidats opérée par l'ensemble de l'examen d'entrée.

1.1. Propositions stratégiques

1- Les examinateurs doivent être formés à la docimologie. Un appel à candidature pour une formation anticipée doit être formulé de façon à ce que l'effet de halo, ou d'autres biais d'évaluations puissent être débattus, afin de clarifier la position des examinateurs. Dans ce cadre il sera important de préciser que l'effet de contraste est important (c'est un concours) mais qu'il ne doit pas être confondu avec l'effet d'ordre.

2- Faire prendre conscience au jury qu'il faut toujours garder à l'esprit que la pertinence des questions reste discutable. Il s'agit en effet d'un travail virtuel, dans le sens où elles sont évaluées en dehors du contexte d'une situation de travail professionnelle. La figure suivante peut être un outil qui nous semble particulièrement didactique lorsque nous la restituons dans son contexte. Les auteurs précisent l'image en expliquant qu'il s'agit d'un trappeur devant s'appliquer à répondre à deux commandes destinées à évaluer sa compétence à chasser le gibier. Tout d'abord, il lui est demandé de tirer sur une cible classique, ce qui permet de valider sa performance en termes de fiabilité et de validité. Dans un second temps, les tirs s'appliquent à une cible mouvante représentant fidèlement un animal réel. Les résultats obtenus reflètent alors la pertinence de son habileté et la connaissance de l'habitus de l'animal dans ses déplacements, donc de sa compétence à agir

en situation réelle. La première série de tirs peut être comparée à l'épreuve écrite d'un examen qui permet d'évaluer la validité et la fiabilité ces connaissances d'un candidat.

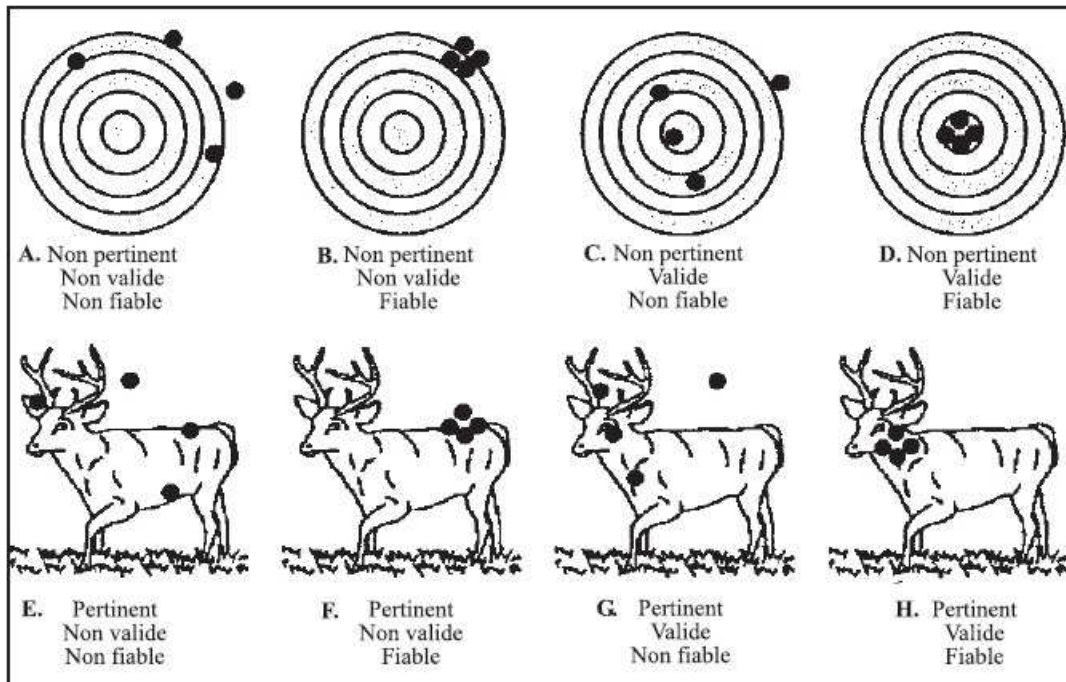


Figure 33 : Pertinence, validité et fiabilité (D'après DE KETELE et GERARD (2004))

Les auteurs de cette figure expliquent les différents éléments qui la compose de la façon suivante :

« Dans la situation A, le tir est non fiable parce que dispersé sur toute la surface de la cible fixe. Il est non valide, parce que la mesure manque de précision et ne peut attester l'atteinte de l'objectif. Il est enfin non pertinent puisqu'il ne correspond pas à l'objectif recherché ...

Dans la situation B, le tir est groupé mais rate systématiquement la cible. Il est fiable (la mesure montre que le tireur sait faire un tir groupé), mais non valide parce que cette mesure ne me permet pas de vérifier qu'il est capable d'atteindre la cible.

Dans la situation C, le centre de la cible, c'est-à-dire l'objectif, est atteint et la mesure est donc valide, mais elle n'est pas fiable parce qu'on ne peut certifier que ce sera toujours le cas.

Dans la situation D, le tir est groupé et touche le mille, mais la cible n'est pas la bonne (par rapport à la compétence visée). La mesure est donc fiable et valide, mais non pertinente.

Les situations E à H présentent des situations pertinentes par rapport à la compétence visée. Les situations E, F et G ont cependant des problèmes de validité

*et/ou de fiabilité, alors que la situation H détermine une mesure à la fois fiable, valide et pertinente. Elle seule permet de certifier que le trappeur est compétent. Ces derniers mots sont importants: l'évaluation a pour objectif de certifier que le trappeur est compétent pour chasser du gibier.*¹³²»

Cette notion de pertinence est essentielle. Elle fait partie des notions que tous les membres du jury doivent garder à l'esprit.

3- Généraliser la pratique des préparations au concours d'entrée pour tous les candidats. Comme nous l'avons dit lors de notre présentation des composantes humaines de l'examen¹³³, la grande majorité des premiers admis ont bénéficiés de ce type de formation. L'objectif de celles-ci n'est pas d'améliorer les connaissances et les compétences professionnelles du candidat, mais de les préparer à savoir les mettre en valeur aussi bien à l'écrit qu'à l'oral. Il existe donc un déficit de chances pour les personnes qui ne sont pas préparées, alors qu'elles possèdent peut être toutes les qualités requises pour suivre la formation d'Infirmier Anesthésiste. Par ailleurs, ces sessions de préparation coutent cher et ne sont pas toujours prises en charge par les établissements employeurs, elles devraient donc être proposée de préférence gratuitement aux postulants. Ceci est d'autant plus possible que ces formations font partie des missions règlementaires des écoles d'IADE. Notons également qu'il n'a jamais été stipulé qu'elle devait être rémunérée. Nous considérons qu'il s'agit d'un travail pour la fonction publique hospitalière, et au delà pour toute la population. Il n'est en effet pas question de l'intérêt individuel de la personne qui postule, mais plutôt de l'intérêt général.

Par ailleurs nous avons pu mettre en évidence que l'épreuve orale, fonctionne bien. Sa fonction de sélection est efficace, comme en atteste sa courbe de distribution bimodale des notes attribuées lors des entretiens. Nous notons que celle-ci est en accord avec le résultat final, qui lui-même est prédictif des résultats à l'issue des deux ans de la formation. En conclusion, nous dirons « qu'on ne change pas une équipe qui gagne » et, comme le disait une personne qui m'est chère¹³⁴ « Quand ça marche, surtout n'y touche pas »¹³⁵

¹³² De Ketele JM, Gerard FM. (2005): «*La validation des épreuves d'évaluation selon l'approche par les compétences* »

¹³³ P. 62

¹³⁴ Pierre PELLETIER (1911-1979) « dit Maurice », mécanicien et metteur au point émérite : mon grand père

¹³⁵ En référence à N. DENOIEL « *Le biais du gars* ».

1.2. Propositions d'ingénierie de conception

1- Nous devons prévoir une réunion préparatoire avec les membres du jury en amont de celle destinée au choix des questions que nous avons évoquée dans la seconde partie de ce travail¹³⁶. Celle-ci aura pour objectif de présenter les caractéristiques techniques des questions écrites susceptibles de leur attribuer un pouvoir discriminant suffisant. Ainsi, tous ceux qui le souhaitent pourront proposer des questions techniquement compatibles avec la finalité de sélection de l'examen. Dans ce cadre, les résultats de notre étude concernant les caractéristiques et la typologie des questions les plus discriminantes seront présentés, de manière à apporter une méthodologie de réalisation des items. Nous mettrons en avant les éléments suivant :

- Les Questionnaires à Choix Multiples et les Questionnaires à Réponses Ouvertes Courtes sont à privilégier au regard de leur pouvoir discriminant à priori supérieur aux autres types de questions.
- Parmi les QCM, les questions traitant de la transfusion sanguine sont particulièrement intéressantes, tant sur le plan de leur importance professionnelle que sur celui de leur indice de discrimination élevé, en effet, cent pour cent des questions de ce type proposées présentent les caractéristiques d'items discriminants. Par ailleurs, la rédaction d'un QCM généraliste est à plus de soixante dix pour cent prédictif d'une question discriminante. Ceci est particulièrement intéressant en termes de fiabilité de correction et de diversité des choix de sujets à aborder.
- Si des questions de type « Vrai – Faux » sont proposées, un barème de notation spécifique devra être appliqué, en demandant un nombre minimum de réponses justes à partir duquel des points seront attribués, de façon à limiter la pratique du « guessing ». Dans tous les cas, il faudra insister sur la limite du pouvoir discriminant de ces questions qui n'est suffisant que pour une question sur quatre dans notre étude.
- La rédaction des QROC doit être réservée aux schémas anatomiques à légender ainsi qu'aux questions de calculs qui ne réclament que le résultat. Si le détail d'un raisonnement mathématique est demandé, l'item sera classé dans les Questionnaires à Réponses Ouverte de longueur Moyenne, donc plus sujettes à interprétation.

¹³⁶ P. 68

- La rédaction de Questionnaires à Réponses Ouverte de longueur Moyenne est possible, à la condition d'encadrer le volume de mots destinée à la réponse afin de ne pas encourager des réponses rédactionnelles plus difficiles à coter. En effet, nous serions alors placés dans le cadre des Questions à Réponses Ouvertes Longues dont la typologie n'a pas été relevée dans notre travail et dont on sait qu'elle induisent plus de biais de correction.
- Il faudra enfin insister sur le fait qu'un item présentant beaucoup de difficulté pour les candidats, n'est en aucune façon à considérer de fait comme discriminant.

2- Chaque année, un bilan, faisant état des indices D de discrimination et P de difficulté calculés pour chaque item proposé, doit être réalisé afin de pouvoir continuer à alimenter une base de données de questions étayée qualitativement. Celle-ci permettra une construction mieux maîtrisée des épreuves de sélections. Il sera bien entendu nécessaire d'adapter le contenu de ces items et les réponses attendues, de manière à varier les sujets abordés, tout en conservant leur forme de construction initiale.

3- Les propositions faites par les membres du jury devront être soumises à une personne compétente en la matière, ou à un comité formé à cela, en charge de trier et proposer à l'ensemble du jury contre argumentation, une série de 20 questions choisies pour leur caractère discriminant et la nécessaire diversité des sujets abordés

4- A cette occasion, un dosage adapté de ces questions devra être appliqué pour l'épreuve considérée. Les résultats de notre étude sur le sujet nous amène à proposer d'insérer à hauteur d'au moins 55 % (soit au moins onze questions sur vingt) des items dont l'indice de discrimination est supérieur à 0,4¹³⁷. Parmi ceux-ci, environ 70 % (soit sept ou huit sur onze dans l'exemple précédent) devront avoir un indice de discrimination supérieur à 0,5¹³⁸, dans la mesure où ils auront pu être testées ou évaluées au préalable. Ceci implique bien sûr une marge d'incertitude liée à la proposition de nouvelles questions.

5- Il nous faudra prendre en compte la possibilité d'une utilisation des moyens informatiques dédiés à l'instrumentation de la correction des Questions à Choix Multiples ou des « Vrai – faux », ayant pour objectif de limiter les erreurs de correction. Ceci existe déjà dans les examens de masses tels les concours d'entrée aux Instituts de Formation en Soins Infirmier ou les Examens Nationaux Classant permettant l'accès à l'internat de médecine.

¹³⁷ P. 90

¹³⁸ P. 91

6- Enfin, l'indice P sera utilisé pour adapter le niveau de difficulté de l'épreuve écrite au nombre de candidats inscrits aux épreuves de sélection, afin de limiter le volume de personnes admissibles à l'épreuve orale. Cette proposition d'ingénierie possède également une dimension stratégique pour l'établissement de formation qui doit composer avec des ressources matérielles assez limitées, en termes de quantité et de dimensions des locaux par exemple, ainsi que la difficulté de disposer d'un nombre suffisant de membres compétents et disponibles pour composer le jury.

Par ailleurs une communication avec les examinateurs sera nécessaire pour éviter la note de 20/ 40 à l'épreuve orale, rare soit, mais qui peut être encore améliorée. Le positionnement de ces professionnels est essentiel dans leur choix, et doit faire l'objet de débats clairement argumenté lors d'une harmonisation des notes entre les deux jurys. Cette attitude devrait permettre de gommer un peu plus le nombre de postulants autour de la moyenne¹³⁹.

2. REINVESTISSEMENT PROFESSIONNEL DE L'ETUDE

Le réinvestissement professionnel de notre de travail est clairement de l'ordre de l'ingénierie. D'une part il permet de fiabiliser la sélection des candidats admissibles à l'épreuve orale, donc la sélection finale, et d'autre part il évite les frustrations des candidats liées à l'attribution d'une note trop proche de la limite d'admissibilité.

La mise en œuvre des ces propositions va nous obliger à concevoir un nouveau dispositif de fonctionnement. Celui-ci va nécessiter une organisation d'amont différente, axée sur la formation des membres du jury, ce qui sollicitera un peu plus nos partenaires. Le challenge risque d'être difficile lorsqu'on connaît les possibilités de disponibilité de ceux-ci, déjà très occupés sur les terrains de la clinique et de la recherche.

En revanche, nous pouvons souhaiter que les probabilités de litiges d'ordre règlementaires se trouveront de fait plus limités par une sélection claire des candidats à l'entrée en formation.

¹³⁹ P.75

3. CONCLUSION GENERALE

Rédiger la conclusion d'un travail c'est un peu comme dire au revoir. C'est une tâche délicate qui doit permettre à l'auteur de s'en aller discrètement sans laisser aux lecteurs un sentiment d'abandon.

Nous garderons de ces moments de travail le souvenir intense d'une recherche de connaissance et de solutions espérées. L'approfondissement de la science des examens fut pour nous une source d'apports nouveaux et complémentaires grâce à l'angle de vue différent procuré par la démarche de recherche. Nous avons la prétention de penser qu'à travers elle, nous avons pu développer des compétences nouvelles qui nous seront utiles dans les mois et les années qui viennent. Au delà de ceci, nous pensons que le choix de ce sujet n'était pas le fruit du hasard, mais plutôt celui d'un sentiment d'injustice, souvent vécu dans notre enfance, face aux résultats quelques fois fort décevants au regard de l'enthousiasme développé pour la réalisation du travail demandé. Pour cela, nous espérons que ces compétences acquises pourront être réinvesties dans notre pratique professionnelle quotidienne, pour ne pas reproduire chez ceux qui sont maintenant nos étudiants, le même sentiment d'abandon et d'incompréhension qui fut naguère le notre.

Certains constats furent particulièrement réconfortants, nous pensons en particulier à l'étude préliminaire des résultats au concours d'admission par rapport aux résultats à l'issue de formation qui laissent à penser que notre ouvrage quotidien n'est pas dénué d'efficacité. C'est une chaleur toujours bénéfique au cœur d'un formateur.

Ce travail n'est pas terminé, il va nous demander, dans les années qui viennent, de vérifier la validité des hypothèses avancées au cours de cette étude. Nous aborderons ainsi la phase d'expérimentation qui nous paraît indispensable à la complétude de la recherche et à la confirmation que tout cela n'est pas vain.

Les outils méthodologiques que nous avons développés sont d'ores et déjà opérationnels et nous comptons les réutiliser très rapidement tel que nous l'avons proposé, pour notre propre compte, mais aussi auprès d'autres établissements de formation qui pourront nous en faire la demande. Ceci doit concourir à la démarche permanente d'amélioration de la qualité dans les organismes de formations professionnelles en santé et plus particulièrement dans les Ecoles d'Infirmiers Anesthésistes Diplômés d'Etat.

Nous allons donc prendre congé, tout en gardant à l'esprit que le souvenir de ces quelques mois passé dans l'enceinte de l'Université François RABELAIS nous accompagnera

chaque fois que nous ferons appel à nos compétences en Stratégie et Ingénierie en Formation d'Adultes.

Références et Index

TABLE DES FIGURES

Figure 1 : Ecart des notes entre correcteurs. D'après H. PIERON	27
Figure 2: Nombre de correcteur pour une "vraie note" D'après H. PIERON	28
Figure 3: exemple de boîte de SKINNER	34
Figure 4 : candidats reçus ou refusés (selon H. PIERON 1963).....	46
Figure 5: Rang au Diplôme d'Etat / Rang relatif au concours (sur 5 ans)	49
Figure 6: Rang au Diplôme d'Etat / Rang relatif au concours et zone d'incertitude	50
Figure 7: droite de régression et corrélation	51
Figure 8: résultats du test de SPEARMAN	53
Figure 9: extrait de la table de SPEARMAN.....	53
Figure 10: allure caractéristique d'une courbe de GAUSS	58
Figure 11: Distribution théorique des candidats au concours.....	59
Figure 12: Distribution souhaitée à l'issue du concours	60
Figure 13: répartition à l'arrivée en formation (A) et répartition souhaitée à la fin des études (B)	60
Figure 14: Distribution des notes finales de l'examen	71
Figure 15: Répartition des notes d'écrits sur 7 ans	72
Figure 16: Répartition des notes d'écrits sur 7 ans (Intervalle réduit).....	73
Figure 17: corrélation des notes d'écrits par rapport aux notes finales.....	74
Figure 18: distribution des notes de l'épreuve orale sur 6 ans	75
Figure 19: corrélation des notes d'oraux par rapport aux notes finales	77
Figure 20: corrélation entre l'épreuve écrite et l'épreuve orale.....	78
Figure 21: interprétation des valeurs de l'indice D (EBEL-1965)	83

Figure 22: exemple de résultats discrimination et difficulté.....	87
Figure 23: Exemple de diagramme polaire discrimination et difficulté	88
Figure 24: Corrélation indice D / indice de difficulté.....	89
Figure 25: Nombre de questions discriminantes par année	90
Figure 26: Valeur de D / intervalle autour de la moyenne.....	90
Figure 27: test de SPEARMAN : indice D Vs % notes entre 9,5 et 10,5	91
Figure 28: Indices D et P des questions 2007	92
Figure 29: Représentation graphique des Indices D et P des 20 questions 2007.....	93
Figure 30: Indices D et P des questions 2010.....	94
Figure 31: Représentation graphique des Indices D et P des 20 questions 2010.....	95
Figure 32: Tableau récapitulatif des questions discriminantes	103
Figure 33 : Pertinence, <i>validité et fiabilité</i> (D'après DE KETELE et GERARD (2004))...	108

REFERENCES BIBLIOGRAPHIQUES

- 1 Antibi A. (2003) : « *La Constante macabre* », Toulouse : Ed. Nathan Col. Math'Adore, 159 p.
- 2 Arendt H. (2005). « Responsabilité et jugement », Paris :Ed.Payot,.Coll. Essais. 320p.
- 3 Aubégnay J. (1987) : « *Les pièges de l'évaluation* », Maurecourt, Ed. Editions Universitaires UNMFREO, 266 p.
- 4 Aubert N. et col (1991). « *Management aspects humains et organisationnels* » : Paris : Ed. PUF ; Col. Fondamental. 654 p.
- 5 Barbier R. (1996). « *La recherche-action* » Éd. ECONOMICA, Coll. Anthropos 109 p.
- 6 Bonboir A. (1972). « *La docimologie* » : Paris : Ed. PUF Col. Sup. 195 p.
- 7 Bridgman C.S. (1964) : « *The relation of the upper-lower item discrimination index, D, to bivariate normal correlation coefficient.* » In Educational and Psychological Measurement Vol. XXIV, n°1. p. 85

- 8 Capele C. (2010) : « *Pratiques de correction sur copies d'examen et nouveaux usages instrumentés* » Doctorante CIFRE NEOPTEC, Sous la Direction de Chantal Charnet. PRAXILING UMR 5267 Université MONTPELLIER III – CNRS. En ligne : <http://www.educpros.fr/>
- 9 Cardinet J., et Tourneur Y. (1985) : « *Assurer la mesure* ». Berne - Francfort-s.Main - New York: Peter Lang.
- 10 Carré P. et Caspar P. (2004). « *Traité des sciences et techniques de la Formation* » : Paris : Ed : Dunod (2^{ème} édition). 600 p.
- 11 Carré P., Moisan A., Poisson D. (1997). « *L'autoformation : psychopédagogie, ingénierie, sociologie* ». Paris : PUF. (2^{ème} éd. 2002). 276 p.
- 12 Caverni JP. (1981) : « *La fiabilité de la notation scolaire par expert comme fonction du mode d'exploitation analytique ou synthétique de la note* ». In: L'année psychologique. 1981 vol. 81, n°2. pp. 369-384.
- 13 Chrétien F. (2006) : « *Etude docimologique des questions à choix multiples de l'épreuve d'histologie du concours de PCEM1 à la faculté de médecine de Créteil/Paris XII* » Mémoire DIU de Pédagogie Médicale.
- 14 Cibois Ph. (2003). « *Les écarts à l'indépendance. Techniques simples pour analyser des données d'enquête.* ». In *Sciences Humaines* 2003.
- 15 Crozier, M., Friedberg, E. (1977). « *L'acteur et le système* ». Paris : Ed :Seuil. 436 p
- 16 Dague P. (1952). « *Le régime des examens et des concours en France : problèmes de docimologie* ». In: *Enfance*. Numéro Spécial, 1952. pp. 447-471.
- 17 Dauvisis MC. (1999) : « *Valeurs latentes des pratiques d'évaluation certificatives* » in *Spirales-Revue de Recherches en Education*, 1999 n°23, pp231-241.
- 18 De Ketele JM, Gerard FM. (2005): «*La validation des épreuves d'évaluation selon l'approche par les compétences* »: In : *Mesure et évaluation en éducation*, 2005, Vol. 28, N° 3, pp 1-26.
- 19 De Ketele JM. (1993) : « *L'évaluation conjuguée en paradigmes* ». In: *Revue française de pédagogie*. Volume 103, 1993. pp. 59-80
- 20 De Landsheere G. (1971). « *Evaluation continue et examens. Précis de docimologie* » : Paris : Ed. LABOR Fernand Nathan ; Ed. Education 2000. 238p.

- 21 De Landsheere G. (1979). « *Dictionnaire de l'évaluation et de la recherche en éducation* » : Paris : Ed. PUF (1^{ère} édition). 338 p.
- 22 De Landsheere V. (1971). « *Faire réussir, faire échouer : la compétence minimale et son évaluation* » : Paris : Ed. PUF ; Col. Pédagogies d'aujourd'hui. 255 p.
- 23 De Péretti A. et col. (2000). « *Encyclopédie de l'évaluation en formation et en éducation* » : Paris : Ed. ESF Editeur ; Col : Pédagogie outils. 555 p.
- 24 De Singly F (2011) : « *L'enquête et ses méthodes : Le questionnaire* », Paris, Ed. Armand Collin, Col. 128 ; (2^{ème} édition) ; 127 p.
- 25 Degenne A. (2001) « *Introduction à l'analyse des données longitudinales* » In *Sciences Humaine* 2003.
- 26 Delzongle F. et Gérard J. (2008). « *Illusion de la mesure, réalité de la communication* » : in Cahiers Pédagogiques n°438 (dec. 2005) pp 19-21.
- 27 Desrosières A. (2001). « *Entre réalisme métrologique et conventions d'équivalence : Les ambiguïtés de la sociologie quantitative* » in *Genèse* N° 43, pp 112-127.
- 28 Didier C. (1997). « *S'accorder sur ce qui importe, la pluralité du jugement sur les compétences dans la formation initiale en soins infirmiers.* ». Mémoire : Maitrise de Sciences et Techniques de gestion et animation des systèmes de formation ; Université Paris Dauphine.
- 29 Dixmier F. (2010). « *La vraie fausse note* » ; Pénombre – La lettre grise – www.olivier.hammam.free.fr/mports/penombre/lg7/lg07b.htm
- 30 Durand D. (2004). « *La systémique* ». Paris : Ed.PUF (9^{ème} édition). 126 p.
- 31 Durand-Colson E. (2000) : « *Rapport de l'Inspection Générale de Philosophie sur les notes et les sujets de philosophie au baccalauréat.* » : Rapport IGEN.
- 32 Ebel R.H. (1965) : « *Measuring educational achievement* ». Ed. Prentice Hall, pp. 348-349.
- 33 Esnard C. (2009) : « *Le jugement social* ». Paris : Ed. Dunod. Col. Les Topos. 124 p.
- 34 Findley W. G. (1956). « *A rationale for evaluation of item discrimination statistics* » In. *Educational and Psychological Measurement*, n°16, pp.175-180.
- 35 Grawitz, M. (1993). « *Méthodes des sciences sociales* ». Paris : Dalloz. (9^{ème} éd.). 870 p.
- 36 Grémy J-P (2003). « *Introduction à la lecture des tableaux statistiques* » In *Sciences Humaines* 2003.

- 37 Grenier B., Dubreuil M., Journois D. (2000). « *Comparaison de deux méthodes de mesure d'une même grandeur : méthode de Bland et Altman* » In Ann Fr Anesth Réanim 2000 ; 19 : Ed ; Elsevier, Paris. pp 128-35.
- 38 Groot G. (1972) : « *Le problème de l'évaluation dans l'enseignement des langues vivantes* » in cycle de linguistique appliquée à la pédagogie des langues, Vol. 3. Conférence de Hasselt 1972
- 39 Hadji C. (1989) « *L'évaluation règles du jeu : des intentions aux outils* » Paris : ESF Editeur, Coll. Pédagogies, 190 p.
- 40 Hérant F. (1984). « *L'assise statistique en sociologie* » in « *Economie et statistiques* » N° 168, Juillet-Août 1984, pp 23-35.
- 41 Houssaye J. (2000) : « *Le triangle pédagogique - Théorie et pratiques de l'éducation scolaire* » volume 1, Paris : Ed. Peter Lang, 300p.
- 42 Jodelet, D. (1991). « *Les représentations sociales* » .Paris : PUF. 424 p.
- 43 Kelley, T.L. (1939). « *The selection of upper and lower groups for the validation of test items* ». in The Journal of Educational Psychology, n°30, pp. 17-24.
- 44 Lapointe J. (1993) : « *L'approche systémique et la technologie de l'éducation* » in Educatechnologies (revue en ligne) : Quebec : Université Laval, Vol. 1, n°1 (fév. 1993)
- 45 Le Boterf G.(2003). « *Développer la compétence des professionnels* » : Paris : Ed. Editions d'organisation (4^{ème} édition). 311 p.
- 46 Le Guen M. (2002). « *La boîte à moustache pour sensibiliser à la statistique* ». In *Bulletin de méthodologie sociologique*, 73 (2002) pp 43-64.
- 47 Leclercq D. (2005) « *Edumétrie et docimologie pour praticiens chercheurs* ». Editions de l'Université de Liège – Chapitre 4 « Une Taxonomie des Objectifs Instrumentée au Service de l'Evaluation » (TOISE)
- 48 Leclercq D., Nicaise J., Demeuse M. (2004) : « *Docimologie critique : Des difficultés à noter les copies et d'attribuer des notes aux élèves* ». in : « *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* ». Liège : Ed.Les Editions de l'Université de Liège.
- 49 Leclercq, D. (2006) : « *L'évolution des QCM* ». In G. Figari et L. Mottier-Lopez :« *Recherches sur l'évaluation en Education* ». Paris : L'Harmattan, pp 139-146.

- 50 LOI no2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires
- 51 Martin J. (2002) : « *Aux origines de la « science des examens » (1920-1940)* », Histoire de l'éducation n°94 année 2002, mis en ligne le 08 janvier 2009. URL : <http://histoire-education.revues.org/index817.html>
- 52 Merle P. (2000) : « *Pas de précipitations* » In Cahiers Pédagogiques n°387 (oct. 2000) pp 47-48.
- 53 Merle P. (1998) : « *L'évaluation des élèves, enquête sur le jugement professoral : »* : Paris : Ed PUF, 1996 : in : Revue française de pédagogie, Année 1998, Volume 122, Numéro 1 pp. 181 – 183.
- 54 Mion G. (2009). « *Eléments indispensables de statistiques médicales. Partie 4 - Méthodes non paramétriques* » In *Urgence Pratique* - 2009 N°93.
- 55 Mion G., Herault S., Libert N., Journois D. (2010). « *Eléments indispensables de statistiques médicales. Opus 5 - Analyses multivariées - régression multiple – régression logistique – ACP* » In *Urgence Pratique* - 2010 N°102.
- 56 Mucchielli A.(2009) : « *Dictionnaire de méthodes qualitatives en sciences humaines* ». Paris : Ed. Armand Colin ; (3^{ème} édition), 303p.
- 57 Normand S. Cours : « *L'analyse d'items, une technique docimologique pour valider des questions d'examen* ». Université de Montréal ; Faculté de médecine ; Bureau d'évaluation ; Docimologie appliquée aux sciences de la santé.
- 58 Paicheler G. (1995) : « *Présentation. Les professions de soins : territoires et empiétements* ». In: Sciences sociales et santé. Volume 13, n°3, 1995. pp. 5-10.
- 59 Perrenoud Ph. (2001) « *Développer la pratique réflexive dans le métier d'enseignant. Professionnalisation et raison pédagogique* » Paris, ESF 219 p.
- 60 Picot G. (2005) : « *Entre médecins et personnel infirmier à l'hôpital public : un rapport social instable* » Le cas de deux services hospitaliers, Revue Française des Affaires sociales, 2005/1 n°1, p. 83-100.
- 61 Piéron H. (1963). « *Examens et Docimologie* » : Paris : Ed. PUF ; Col. Le psychologue. 190 p.
- 62 Plenchette M. (1992) : « *Les entretiens d'admission : méthode et stratégies* » ; in Soins

Formation – Pédagogie – Encadrement n°2 ; 2^{ème} trimestre 1992, pp 47 – 51.

- 63 Quinton A. (1997) : « *Pertinence de l'enseignement du PCEM1 au travers de l'analyse docimologique des épreuves de sélection* ». Colloque National : « *Passé, présent et futur de la sélection et de la formation des médecins* » Paris 22 et 23 mars 1997.
- 64 Quinton A. (2005) : « *Docimologie* ». D. U. de pédagogie des Sciences de la Santé ; Université Victor Segalen, Bordeaux 2.
- 65 Quivy, R., Campenhoudt (Van), L., (2000), « *Manuel de la recherche en sciences sociales* », Paris :Dunod
- 66 Raynal F., Rieunier A. (2005) : « *Pédagogie : dictionnaire des concepts clés* ». Paris : Ed. ESF Editeur (5^{ème} édition). 420 p
- 67 Ricœur P.(1995). « *Éthique et responsabilité* », Neuchâtel : La Baconnière.
- 68 Roelens N. (2003). « *Interactions humaines et rapports de force entre subjectivités* » : Paris : Ed. L'Harmattan ; Col. Logiques sociales. 550 p.
- 69 Rosenthal R., Jacobson L. (1978) : « *Pygmalion à l'école* ». Paris : Ed. Casterman.
- 70 Schön D. (1993) ; « *Le praticien réflexif. À la recherche du savoir caché dans l'agir professionnel* » Montréal, Éd. Logiques,.418 p.
- 71 Schratz M. (2001) : « *L'autoévaluation, une nouvelle mission du système éducatif* » in « *Education et société* », Bruxelles, Ed. De Boek Université N° 8 ; pp 65-80.
- 72 Suchaud B. (2008) : « *La loterie des notes au bac. Un réexamen arbitraire de la notation des élèves* » Les Documents de Travail de L'IREDU (Institut de Recherche sur l'EDUcation) en ligne : <http://www.u-bourgogne.fr.iredu>.
- 73 Therer J. (1999) : « *Evaluer pour évoluer, éléments de docimologie* » : Liège : Université de Liege : Laboratoire d'enseignement multimédia.
- 74 Trahan M. et Dassa C. (1978) : « *La courbe normale en psychométrie et en éduométrie : postulat, constante naturelle ou erreur scientifique?* » Revue des sciences de l'éducation, vol. 4, n° 1, 1978, pp. 73-80.
- 75 Vergès P. (2001). « *L'analyse des données par les graphes de similitude* » In *Sciences Humaines* 2001.
- 76 Vial M. ((1999) : « *Ceci n'est pas un lexique 25 mots pour la recherche en sciences de*

l'éducation » Col. En question ; Cahier n°25, Université de Provence, Ai-Marseille 1.

- 77 Vial M. (1999) : « *Modèles et logiques de l'évaluation* », Colloque international, Ethique et qualité dans l'évaluation, Université de Reims Pol'evaluate, 25 et 26 oct 1999.
- 78 Vial M., Bonniol JJ. (1997) « *Les modèles de l'évaluation, textes fondateurs avec commentaires* ». Bruxelles : Ed. De Boeck, Col. Portefeuille. 238 p.
- 79 Zerling JP. (1975) : « *L'ordinateur un nouveau pédagogue ?* » in *Mélanges pédagogiques* : en ligne :
http://www.univ-nancy2.fr/CRAPEL/publications_melanges_1975.htm

ANNEXES

Annexe 1 : Table des données : Rang à l'issue de Diplôme d'Etat / Rang relatif au concours

Annexe 2 : Table de SPEARMAN

Annexe 3 : Notes sur 6 ans

Annexe 4 : Analyses de répartitions réalisées à l'aide du logiciel PRISM®

Annexe 5 : Analyse des notes d'écrit de 2005 à 2011

Annexe 6 : 22 questions discriminantes

I. Annexe 1

Table des données : Rang à l'issue de Diplôme d'Etat / Rang relatif au concours

Etudiant	Rang concours relatif	Place D.E.
N° 1	4	1
N° 2	6	2
N° 3	5	3
N° 4	3	4
N° 5	2	5
N° 6	12	6
N° 7	9	7
N° 8	11	8
N° 9	1	9
N° 10	13	10
N° 11	8	11
N° 12	10	12
N° 13	7	13
N° 14	13	1
N° 15	3	2
N° 16	9	3
N° 17	11	4
N° 18	3	5
N° 19	8	6
N° 20	1	7
N° 21	5	8
N° 22	6	9
N° 23	1	10
N° 24	12	11
N° 25	7	12
N° 26	10	13
N° 27	6	1
N° 28	2	2
N° 29	11	3
N° 30	7	4
N° 31	9	5
N° 32	3	6
N° 33	5	7
N° 34	10	8

Etudiant	Rang concours relatif	Place D.E.
N° 35	4	9
N° 36	12	10
N° 37	1	11
N° 38	7	12
N° 39	12	13
N° 40	14	14
N° 41	2	1
N° 42	6	2
N° 43	6	3
N° 44	8	4
N° 45	10	5
N° 46	2	6
N° 47	4	7
N° 48	1	8
N° 49	9	9
N° 50	10	10
N° 51	12	11
N° 52	4	12
N° 53	14	13
N° 54	12	14
N° 55	3	1
N° 56	1	2
N° 57	1	3
N° 58	12	4
N° 59	4	5
N° 60	14	6
N° 61	9	7
N° 62	9	8
N° 63	11	9
N° 64	6	10
N° 65	12	11
N° 66	4	12
N° 67	8	13
N° 68	7	14

II. Annexe 2

Table de SPEARMAN

n	$\alpha = 0.05$	$\alpha = 0.01$		n	$\alpha = 0.05$	$\alpha = 0.01$
5	1.000			31	.356	.459
6	.886	1.000		32	.350	.452
7	.786	.929		33	.345	.446
8	.738	.881		34	.340	.439
9	.700	.833		35	.335	.433
10	.648	.794		36	.330	.427
				37	.325	.421
11	.618	.755		38	.321	.415
12	.587	.727		39	.317	.410
13	.560	.703		40	.313	.405
14	.538	.675				
15	.521	.654		41	.309	.400
16	.503	.635		42	.305	.395
17	.485	.615		43	.301	.391
18	.472	.600		44	.298	.386
19	.460	.584		45	.294	.382
20	.447	.570		46	.291	.378
				47	.288	.374
21	.435	.556		48	.285	.370
22	.425	.544		49	.282	.366
23	.415	.532		50	.279	.363
24	.406	.521				
25	.398	.511		60	.255	.331
26	.390	.501		70	.235	.307
27	.382	.491		80	.220	.287
28	.375	.483		90	.207	.271
29	.368	.475		100	.197	.257
30	.362	.467				

III. Annexe 3

Notes sur 6 ans

2011				2010			
Candidat n°	Ecrit	Oral	Note finale	Candidat n°	Ecrit	Oral	Note finale
1	34,4	39,5	73,9	29	32,8	36	68,8
2	33,2	37	69,8	30	34	33	67
3	32,8	35	68,2	31	31,4	34	65,4
4	30,5	35	65,5	32	33,2	30	63,2
5	30	37	62,2	33	31,8	30	61,8
6	29,6	31	59	34	31	30	61
7	29,4	27	57	35	29,4	30	59,4
8	29,1	30	56,7	36	23	34	57
9	29	29	56,5	37	28,4	27	55,4
10	28,1	26	55,1	38	26,8	28	54,8
11	28	25	54,6	39	26,2	27,5	53,7
12	27,9	25	54	40	25,4	28	53,4
13	27,6	28,5	53,9	41	30,2	23	53,2
14	27,5	25,5	53,4	42	25,4	27,5	52,9
15	26,8	26	52,8	43	28	24,5	52,5
16	26,7	25	52,6	44	31,2	20	51,2
17	26,4	27	52,4	45	24	27	51
18	26,2	27	49,5	46	29,4	20	50,4
19	25,6	23	47,9	47	23,4	26	49,4
20	25,4	21	47,2	48	25,2	22	47,2
21	25,4	20	45,5	49	25,8	20	45,8
22	25,2	21	43,9	50	25,8	20	45,8
23	24,9	18	44,4	51	24,2	20	44,2
24	24,5	17	42,6	52	28,6	13	41,6
25	22,9	15	44,4	53	30,8	9	39,8
26	22,6	15	43,1	54	20,4	19	39,4
27	22,5	15	37,6	55	22,2	17	39,2
28	20,9	10	30,9	56	20	18	38
				57	21,4	16	37,4
				58	22,6	13	35,6
				59	20,2	9	29,2
				60	23,4	5	28,4

2009				2008			
Candidat n°	Ecrit	Oral	Note finale	Candidat n°	Ecrit	Oral	Note finale
61	33,2	37	70,2	90	30,4	30	60,4
62	27	34	61	91	25,2	35	60,2
63	22,8	37	59,8	92	21,6	38	59,6
64	28,6	30	58,6	93	26,2	30	56,2
65	20,4	35	55,4	94	29,6	26	55,6
66	23,6	31	54,6	95	28,4	27	55,4
67	24,4	30	54,4	96	24,4	30	54,4
68	27	27	54	97	23,2	31	54,2
69	20,8	33	53,8	98	23,8	30	53,8
70	21,8	31	52,8	99	25,6	28	53,6
71	22,6	30	52,6	100	26,4	27	53,4
72	23,8	27	50,8	101	22,8	28	50,8
73	24	26	50	102	25,8	25	50,8
74	28	22	50	103	23,2	27	50,2
75	29,4	20	49,4	104	24,2	26	50,2
76	22,8	26	48,8	105	20,4	28	48,4
77	27,8	20	47,8	106	27,4	21	48,4
78	22,8	25	47,8	107	23,4	24	47,4
79	25,6	20	45,6	108	27,2	20	47,2
80	24	20	44	109	24,8	22	46,8
81	20	22	42	110	26,6	20	46,6
82	21,4	20	41,4	111	24,4	22	46,4
83	30,2	18	48,2	112	26	20	46
84	29	18	47	113	20,6	25	45,6
85	23,2	19	42,2	114	23,2	21	44,2
86	22,6	18	40,6	115	23,6	20	43,6
87	21	18	39	116	23,2	20	43,2
88	20,6	18	38,6	117	22,6	20	42,6
89	23,6	10	33,6	118	20,8	20	40,8
				119	26	13	39
				120	21,8	17	38,8
				121	21	17	38
				122	22,8	15	37,8
				123	26,8	10	36,8
				124	20,4	15	35,4

2006				2005			
Candidat n°	Ecrit	Oral	Note finale	Candidat n°	Ecrit	Oral	Note finale
125	33,3	27	60,3	150	33	32	65
126	31,9	28	59,9	151	31	33	64
127	30,8	29	59,8	152	30,4	33	63,4
128	29,9	30	59,9	153	29,8	33	62,8
129	27,8	27	54,8	154	29,8	32	61,8
130	27	26	53	155	29,6	31	60,6
131	26,7	26	52,7	156	28,6	31	59,6
132	26,5	26	52,5	157	27,2	30	57,2
133	26,1	26	52,1	158	27,2	28	55,2
134	26	26	52	159	27	28	55
135	25,9	26	51,9	160	26,8	28	54,8
136	25,5	25	50,5	161	26,6	28	54,6
137	25,4	24	49,4	162	26,6	27	53,6
138	25,3	23,5	48,8	163	26,2	26	52,2
139	25	22	47	164	26	24	50
140	23,9	24	47,9	165	26	23	49
141	23,9	23,5	47,4	166	25,4	23	48,4
142	23,8	23,5	47,3	167	25,4	22	47,4
143	23,5	22	45,5	168	24,8	22	46,8
144	22,2	20	42,2	169	24,6	21	45,6
145	22,2	20	42,2	170	24,2	21	45,2
146	20,5	18	38,5	171	24	20	44
147	20,4	16	36,4	172	24	19	43
148	20,2	15	35,2	173	23,8	18	41,8
149	20	15	35	174	22,6	17	39,6
				175	22,4	15	37,4
				176	22	15	37
				177	22	12	34
				178	21,8	12	33,8

IV. Annexe 4

Analyses de répartitions réalisées à l'aide du logiciel PRISM®

Répartition des notes finales sur 7 ans.

Number of values	271
Minimum	4,500
25% Percentile	10,10
Median	11,95
75% Percentile	13,40
Maximum	17,20
Mean	11,77
Std. Deviation	2,442
Std. Error	0,1483
Lower 95% CI of mean	11,48
Upper 95% CI of mean	12,07
Shapiro-Wilk normality test	
W	0,9912
P value	0,1040
Passed normality test (alpha=0.05)?	Yes

Bin Center	Note/20	% de la pop.
4	0	0,0
5	3	1,1
6	4	1,5
7	6	2,2
8	10	3,7
9	23	8,5
10	33	12,2
11	38	14,0
12	43	15,9
13	45	16,6
14	25	9,2
15	26	9,6
16	10	3,7
17	5	1,8
18	0	0,0

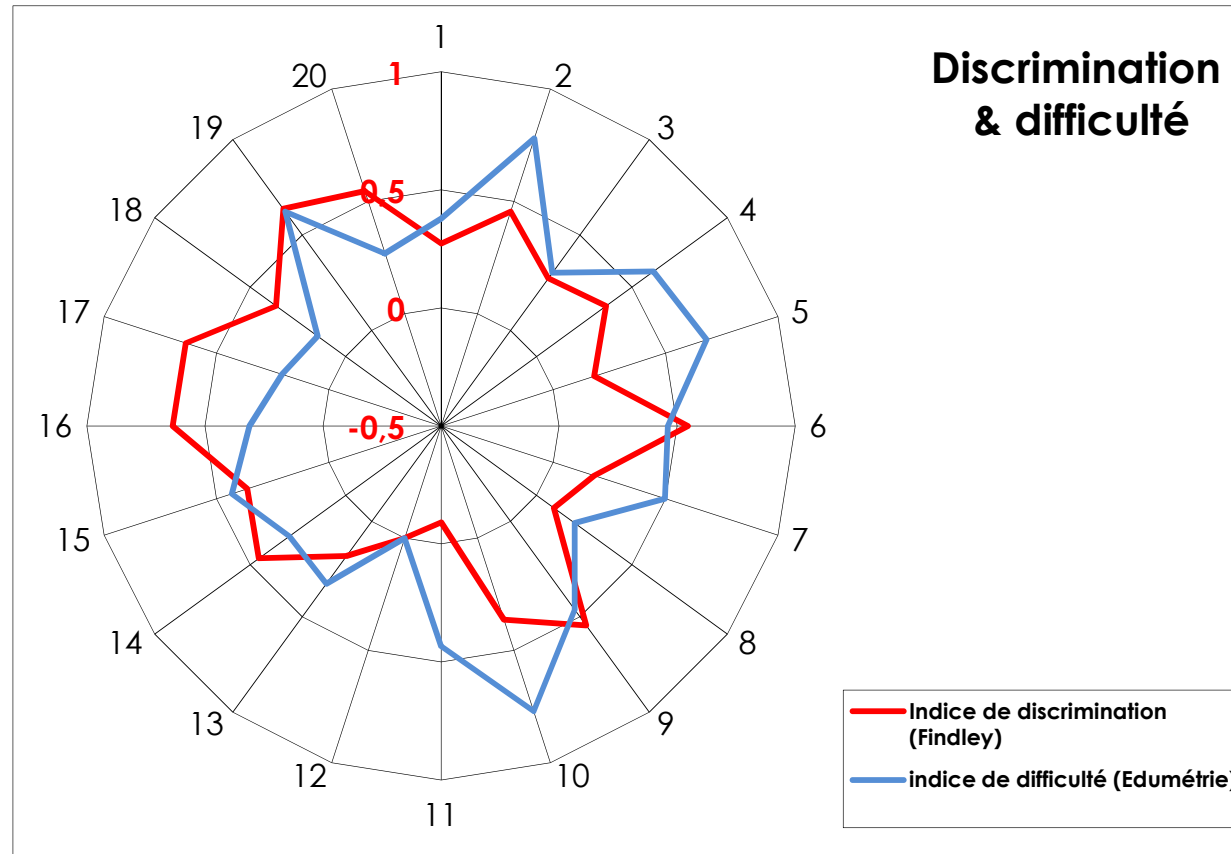
V. Annexe 5

Analyse notes d'écrit 2005

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	10	10	6	8	5	10	6	8	10	0	7	10	9	9	8	9	10	10	10	10	165	33
2	10	0	8	9	3	6	2	10	10	8	7	10	9	9	7	9	10	8	10	10	155	31
3	5	0	8	1	9	10	6	8	10	10	6	10	8	7	10	9	10	10	5	10	152	30,4
4	10	10	4	3	6	9	6	8	10	5	8	10	3	10	9	8	10	10	0	10	149	29,8
5	10	0	4	4	3	8	8	10	10	8	7	10	7	6	6	8	10	10	10	10	149	29,8
6	10	0	10	8	5	9	6	4	10	3	7	10	6	3	8	9	10	10	10	10	148	29,6
7	10	10	6	8	5	8	4	8	0	8	6	10	8	7	5	10	10	10	0	10	143	28,6
8	0	10	8	4	5	9	6	6	10	0	6	10	8	8	6	10	10	10	10	0	136	27,2
9	5	10	8	7	2	4	2	8	0	5	6	10	9	8	6	8	10	8	10	10	136	27,2
10	10	10	6	2	3	4	6	10	10	0	5	10	6	9	7	7	10	10	0	10	135	27
11	5	0	10	2	8	6	8	8	0	0	5	10	9	10	5	8	10	10	10	10	134	26,8
12	5	0	8	7	2	9	4	10	10	8	6	10	5	4	5	5	5	10	10	10	133	26,6
13	5	0	6	7	0	8	4	8	10	0	8	10	5	7	6	9	10	10	10	10	133	26,6
14	5	0	10	6	3	9	2	10	10	10	8	10	7	7	5	8	5	6	0	10	131	26,2
15	0	0	6	4	4	9	4	10	10	0	8	10	8	8	5	9	10	10	5	10	130	26
16	10	0	10	8	4	10	6	6	10	0	8	10	10	10	6	4	10	8	0	0	130	26
17	0	10	10	5	7	0	4	6	0	3	7	10	9	9	8	9	10	10	0	10	127	25,4
18	10	0	4	6	3	9	4	8	0	0	7	10	7	4	7	8	10	10	10	10	127	25,4
19	10	0	6	4	3	8	4	10	5	0	4	10	6	7	5	8	10	4	10	10	124	24,8
20	5	0	6	2	4	7	8	10	10	0	5	10	10	10	6	7	5	8	0	10	123	24,6
21	10	0	10	5	2	5	6	8	10	0	4	10	3	8	6	6	10	8	0	10	121	24,2
22	5	0	8	3	3	0	4	10	0	10	8	10	5	9	3	2	10	10	10	10	120	24
23	5	0	10	3	4	9	4	6	0	3	6	10	9	9	5	7	10	10	0	10	120	24
24	0	10	4	0	4	5	8	8	0	0	5	10	7	6	8	9	10	10	10	5	119	23,8
25	10	0	8	7	3	8	4	10	0	0	9	10	7	9	5	5	5	8	5	0	113	22,6
26	10	0	8	1	5	3	6	10	10	0	8	10	6	8	7	5	5	10	0	0	112	22,4

27	5	0	6	3	2	0	6	10	5	0	6	10	8	9	3	7	10	10	0	10	110	22
28	10	0	10	1	0	6	10	10	10	10	0	10	5	2	6	3	5	2	0	10	110	22
29	10	0	10	3	1	0	6	4	5	0	8	10	5	9	8	5	5	10	0	10	109	21,8
30	10	0	6	1	2	9	6	8	0	0	6	10	5	6	4	7	5	10	0	10	105	21
31	0	0	8	4	3	0	2	8	0	3	6	10	2	8	7	9	10	10	5	10	105	21
32	0	10	4	1	7	5	6	10	0	0	8	10	7	6	7	4	10	10	0	0	105	21
33	10	0	6	3	0	0	4	8	10	0	2	10	8	6	6	6	5	6	0	10	100	20
34	10	0	10	6	1	4	4	10	0	0	8	10	6	0	0	9	10	10	0	0	98	19,6
35	0	0	6	3	2	0	2	8	5	0	7	10	8	8	7	6	5	10	0	10	97	19,4
36	10	0	4	2	3	8	4	0	0	0	2	10	4	4	3	8	5	8	0	10	85	17
37	0	0	8	2	1	0	4	6	5	0	2	10	8	7	3	6	5	10	0	0	77	15,4
38	10	0	4	3	1	0	6	4	5	0	1	10	8	7	4	4	0	4	0	0	71	14,2
39	0	0	4	1	1	3	6	8	10	0	1	10	8	2	3	7	0	2	5	0	71	14,2
40	0	0	4	0	0	0	6	8	0	0	3	10	6	7	4	4	5	8	0	5	70	14
41	5	0	4	2	1	4	2	8	0	0	1	10	2	7	4	1	10	2	0	0	63	12,6

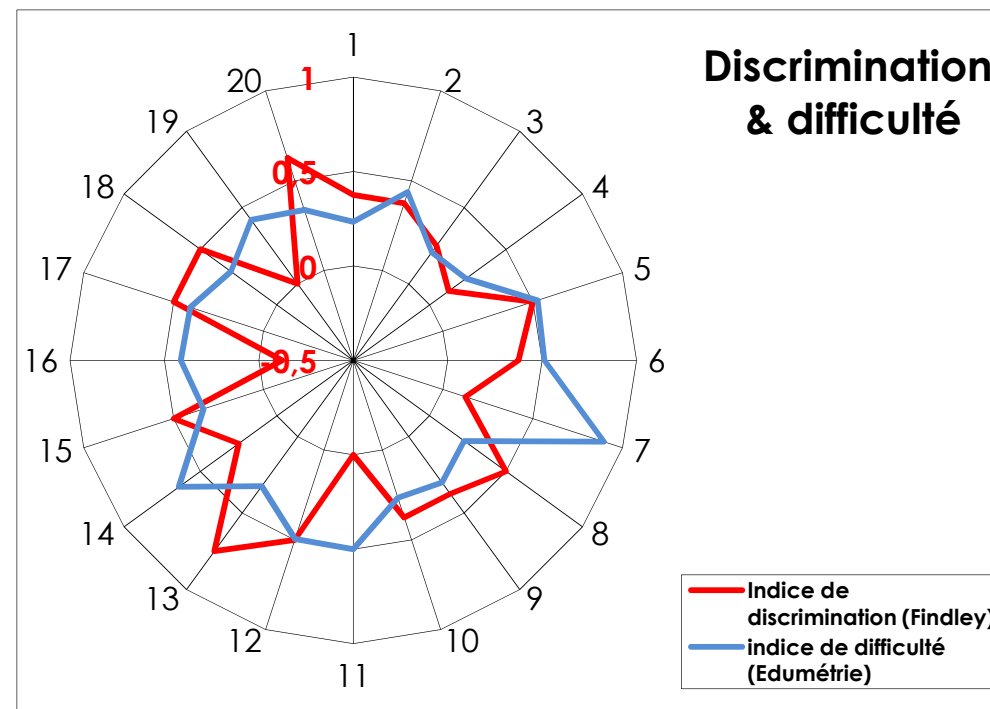
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	6,22	2,2	6,98	3,88	3,17	5,39	5,02	8	5,37	2,29	5,66	10	6,73	7,05	5,68	6,88	7,93	8,54	3,78	7,32
Médiane	5	0	6	3	3	6	6	8	5	0	6	10	7	7	6	7	10	10	0	10
MINI	0	0	4	0	0	0	2	0	0	0	0	10	2	0	0	1	0	2	0	0
MAXI	10	10	10	9	9	10	10	10	10	10	9	10	10	10	10	10	10	10	10	10
Réussite des "FORTS" (>8)	0,64	0,55	0,55	0,36	0,18	0,64	0,18	0,82	0,73	0,36	0,09	1	0,64	0,64	0,36	0,91	1	1	0,64	0,91
Réussite des "FAIBLES"(>8)	0,36	0,09	0,27	0	0	0,09	0	0,73	0,18	0	0,18	1	0,45	0,18	0	0,27	0,36	0,64	0	0,36
Indice de discrimination (Findley)	0,27	0,45	0,27	0,36	0,18	0,55	0,18	0,09	0,55	0,36	-0,09	0,00	0,18	0,45	0,36	0,64	0,64	0,36	0,64	0,55
indice de difficulté (Edumétrie)	0,38	0,78	0,30	0,61	0,68	0,46	0,50	0,20	0,46	0,77	0,43	0,00	0,33	0,30	0,43	0,31	0,21	0,15	0,62	0,27
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,27	0	0,51	0,44	0,51	0,29	0,73	0,2	0,15	0,15	0,39	0	0,37	0,22	0,63	0,32	0,32	0,1	0,12	0,05



Analyse notes d'écrit 2006

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	8	8	10	7	9,5	10	3	9	10	10	5	3,5	9	10	10	10	10	10	4,5	10	167	33,3
2	10	7	9,5	8,5	4	7	0	8,5	9	10	10	10	7	3	10	10	9	10	7	10	160	31,9
3	8	5	10	7	8,5	10	10	9,5	10	10	0	6	9	9	5	2	9	10	7	9	154	30,8
4	8	7	6,5	9	6	7	0	8,5	7	10	10	8	8	6	10	6	7,5	10	7	8	150	29,9
5	8	10	7,5	10	10	10	0	6,5	7	10	0	8	5	3	10	6	4,5	10	5,5	8	139	27,8
6	10	7	9	10	10	6	6	8	7	0	0	8	9	3	5	6	6	10	9	6	135	27
7	8	10	9	7	7	3	0	9	5	10	10	4	8	6	10	6	9	0	5,5	7	134	26,7
8	4	6	9	5,5	5	0	0	8	10	10	10	10	8	7	10	4	7,5	10	4,5	4	133	26,5
9	6	3	10	10	8	7	0	9,5	7	10	10	0	5	0	5	10	7,5	10	4,5	8	131	26,1
10	6	6	8	5,5	5	10	6	7,5	9	10	5	2	7	6	5	2	6	10	7	7	130	26
11	10	7	9	9	6,5	6	3	8,5	10	10	10	8	5	1	5	4	3	0	4,5	10	130	25,9
12	8	6	7	7	7,5	4	0	9	9	10	0	6	10	3	0	6	9	10	8	8	128	25,5
13	10	3	7,5	8	1,5	5	0	7,5	6	10	10	6	7	0	10	6	6	10	4,5	9	127	25,4
14	8	5	9	10	0	7	0	10	5	10	10	4	3	6	10	10	6	0	4,5	9	127	25,3
15	8	4	6,5	6	6	6	3	7	10	0	0	8	8	6	10	6	7,5	10	7	6	125	25
16	10	10	6,5	9	6,5	3	0	7	7	10	0	8	6	3	5	10	6	0	4,5	8	120	23,9
17	6	9	8	7	10	7	0	6,5	7	10	5	4	6	3	5	2	4,5	10	4,5	5	120	23,9
18	10	3	6	9,5	1,5	1	0	8	7	10	0	2	8	6	10	6	3	10	10	8	119	23,8
19	8	9	9	7	8	6	0	7	9	0	0	4	6	3	10	4	6	10	4,5	7	118	23,5
20	6	5	7,5	6	1,5	2	0	7,5	4	10	10	2	5	3	10	6	4,5	10	7	4	111	22,2
21	6	8	7,5	7	10	7	0	7	5	10	0	4	8	0	5	2	3	10	4,5	7	111	22,2
22	10	4	9	6,5	2	6	0	8	6	0	0	4	8	3	10	0	4,5	10	4,5	7	103	20,5
23	8	7	7,5	9	4	0	0	8	7	10	5	4	7	3	5	4	4,5	0	7	2	102	20,4
24	6	1	8	6,5	0	0	0	7,5	5	10	0	4	6	3	10	6	7,5	10	4,5	6	101	20,2
25	4	2	7	7	1	6	0	6,5	4	0	10	8	7	6	10	10	3	0	4,5	4	100	20
26	6	5	4,5	8,5	2,5	6	0	6	7	0	10	2	7	3	5	10	3	0	4,5	3	93	18,6
27	8	4	7,5	5	0,5	0	0	5,5	5	0	10	2	5	3	0	10	6	10	4,5	5	91	18,2
28	8	6	8	10	0	0	0	6,5	5	10	0	4	5	0	0	6	4,5	0	7	6	86	17,2
29	8	1	8	4	0	0	0	7	3	10	10	4	4	0	0	4	3	10	4,5	4	84,5	16,9
30	6	1	6	7	0	4	0	8	8	0	0	4	7	0	0	2	7,5	0	8	4	72,5	14,5

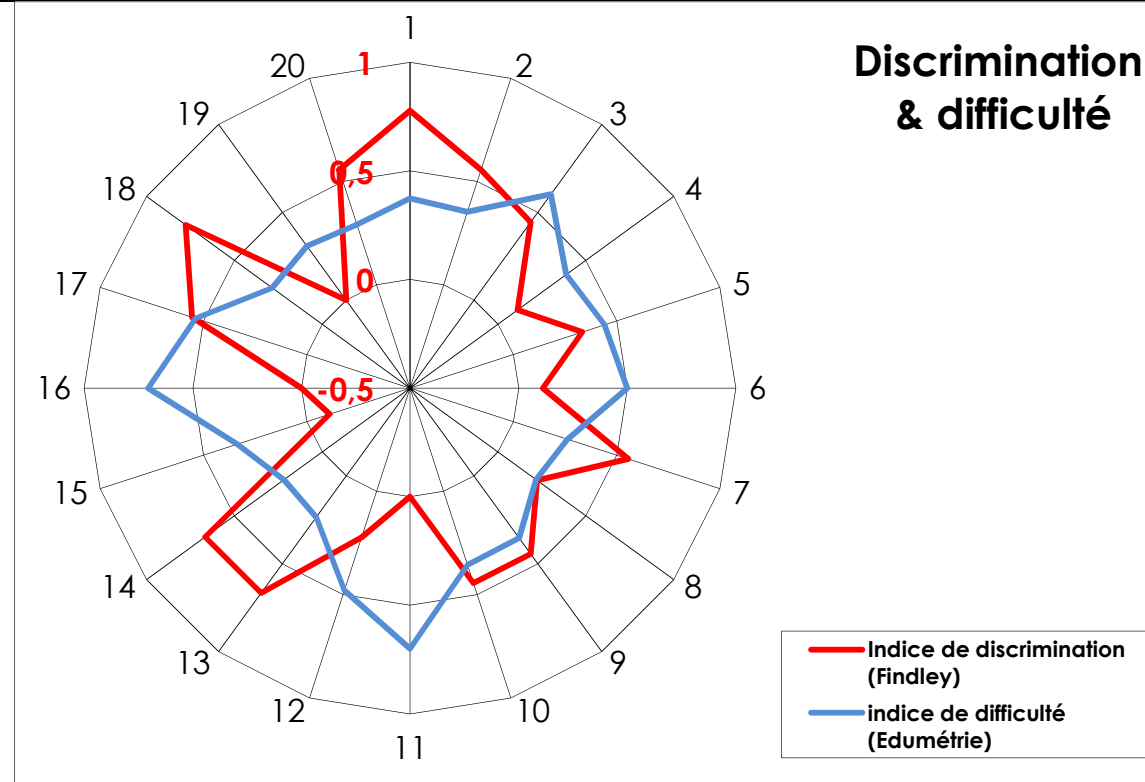
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	7,67	5,63	7,92	7,62	4,73	4,87	1,03	7,73	7	7,33	5	5,05	6,77	3,6	6,67	5,87	5,93	7	5,8	6,63
Médiane	8	6	8	7	5	6	0	7,75	7	10	5	4	7	3	7,5	6	6	10	4,5	7
MINI	4	1	4,5	4	0	0	0	5,5	3	0	0	0	3	0	0	0	3	0	4,5	2
MAXI	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Réussite des "FORTS" (>8)	0,88	0,38	0,88	0,5	0,5	0,38	0,13	0,88	0,5	0,88	0,5	0,63	0,75	0,25	0,75	0,25	0,75	0,88	0,13	0,63
Réussite des "FAIBLES"(>8)	0,5	0	0,63	0,38	0	0	0	0,38	0,13	0,5	0,5	0,13	0	0	0,25	0,38	0,25	0,38	0,13	0
Indice de discrimination (Findley)	0,38	0,38	0,25	0,13	0,50	0,38	0,13	0,50	0,38	0,38	0,00	0,50	0,75	0,25	0,50	0,13	0,50	0,50	0,00	0,63
indice de difficulté (Edumétrie)	0,23	0,44	0,21	0,24	0,53	0,51	0,90	0,23	0,30	0,27	0,50	0,50	0,32	0,64	0,33	0,41	0,41	0,30	0,42	0,34
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,33	0,47	0,2	0,27	0,3	0,4	0,17	0,2	0,37	0	0,13	0,5	0,4	0,67	0,33	0,53	0,63	0	0,6	0,4



Analyse notes d'écrit 2007

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	8	9	5	10	10	4	9	9	10	10	3	8	10	10	10	7	9	8	7	8	164	32,8
2	6	7	10	10	5	6	9	9	10	10	3	6	10	10	9	1	8	10	9	9	157	31,4
3	8	9	10	6	10	6	10	8	10	10	3	5	10	10	6	2	8	10	7	8	156	31,2
4	8	7	5	7	10	6	8	9	10	10	4	9	10	10	5	2	9	10	7	8	154	30,8
5	8	9	10	5	10	4	9	9	10	10	1	7	10	7	6	3	9	9	6	9	151	30,2
6	8	9	5	6	10	8	7	6	10	10	3	4	10	10	8	3	6	10	7	10	150	30
7	6	8	5	6	5	8	9	7	10	2	4	7	10	10	5	3	8	10	8	9	140	28
8	8	6	5	6	10	8	9	7	10	5	4	4	10	10	6	2	6	6	7	9	138	27,6
9	10	9	10	4	0	4	10	9	5	5	5	7	10	10	7	3	2	9	7	8	134	26,8
10	5	8	5	10	5	6	9	8	0	10	5	6	10	10	5	3	3	7	8	8	131	26,2
11	5	6	0	6	5	6	10	9	0	10	4	9	10	10	6	2	6	10	9	5	128	25,6
12	8	5	5	6	10	4	4	8	10	10	2	5	10	7	9	5	3	6	7	4	128	25,6
13	5	8	0	4	10	2	9	6	10	10	6	6	10	10	7	4	0	8	7	6	128	25,6
14	8	8	0	5	0	4	7	8	10	10	5	6	10	7	5	4	5	10	7	7	126	25,2
15	8	6	0	6	5	6	8	10	10	0	5	6	10	10	5	4	3	7	7	9	125	25
16	5	7	5	6	10	2	5	8	10	10	1	2	10	7	5	2	4	9	6	9	123	24,6
17	5	7	0	5	10	4	8	9	10	10	4	6	0	10	9	3	4	7	6	5	122	24,4
18	7	6	5	5	0	6	8	7	0	10	3	4	10	10	9	5	3	8	8	8	122	24,4
19	6	6	5	5	5	4	7	10	0	5	3	6	10	10	10	2	7	8	6	6	121	24,2
20	6	5	5	7	10	8	5	9	0	5	1	4	10	7	5	1	5	9	6	9	117	23,4
21	6	8	5	7	0	8	8	4	10	0	2	6	10	7	8	5	2	5	7	8	116	23,2
22	7	6	0	4	0	6	9	8	10	10	3	4	10	7	7	3	4	10	3	5	116	23,2
23	8	9	5	6	5	2	6	7	0	0	3	4	10	10	5	3	3	9	8	10	113	22,6
24	5	6	0	3	0	6	9	9	5	10	3	4	10	10	7	3	3	7	6	7	113	22,6
25	6	2	5	6	10	6	6	8	0	10	3	6	0	6	9	1	6	7	8	7	112	22,4
26	5	5	0	6	0	8	8	3	10	0	5	3	10	7	9	2	8	7	7	9	112	22,4
27	6	4	5	7	5	2	8	9	0	0	2	6	10	10	5	3	2	4	8	9	105	21
28	5	6	5	5	10	6	5	7	0	0	1	7	0	7	8	1	7	6	7	8	101	20,2
29	5	4	0	7	10	2	5	7	5	10	1	0	10	10	0	5	0	1	5	3	90	18
30	5	4	0	10	0	8	10	10	10	0	4	5	0	0	7	5	1	1	7	0	87	17,4
31	7	8	5	6	0	4	6	8	5	0	1	3	0	0	0	3	6	8	7	9	86	17,2
32	5	5	5	6	0	2	3	8	10	7	2	6	0	0	5	0	6	3	8	5	86	17,2
33	5	4	0	5	10	2	2	6	0	10	1	6	0	0	7	10	2	1	3	6	80	16
34	0	4	5	5	0	2	6	7	10	0	2	0	0	0	9	3	0	1	6	7	67	13,4

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	6,26	6,47	3,97	6,12	5,59	5	7,38	7,82	6,47	6,44	3	5,21	7,65	7,82	6,65	2,94	4,62	7,15	6,91	7,09
Médiane	6	6	5	6	5	6	8	8	10	10	3	6	10	10	7	3	4,5	8	7	8
MINI	0	2	0	3	0	2	2	3	0	0	1	0	0	0	0	0	0	1	3	0
MAXI	10	9	10	10	10	8	10	10	10	10	6	9	10	10	10	7	9	10	9	10
Réussite des "FORTS"(>8)	0,78	0,67	0,44	0,22	0,67	0,33	0,89	0,67	0,89	0,67	0	0,22	1	0,89	0,33	0	0,67	0,89	0,22	1
Réussite des "FAIBLES"(>8)	0	0,11	0	0,11	0,33	0,22	0,33	0,44	0,44	0,22	0	0	0,33	0,22	0,44	0	0,11	0,11	0,22	0,44
Indice de discrimination (Findley)	0,78	0,56	0,44	0,11	0,33	0,11	0,56	0,22	0,44	0,44	0,00	0,22	0,67	0,67	-0,11	0,00	0,56	0,78	0,00	0,56
indice de difficulté (Edumétrie)	0,37	0,35	0,60	0,39	0,44	0,50	0,26	0,22	0,35	0,36	0,70	0,48	0,24	0,22	0,34	0,71	0,54	0,29	0,31	0,29
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,56	0,5	0,56	0,74	0,24	0,56	0,29	0,15	0,12	0,12	0,65	0,71	0	0,03	0,41	0,59	0,5	0,21	0,29	0,24

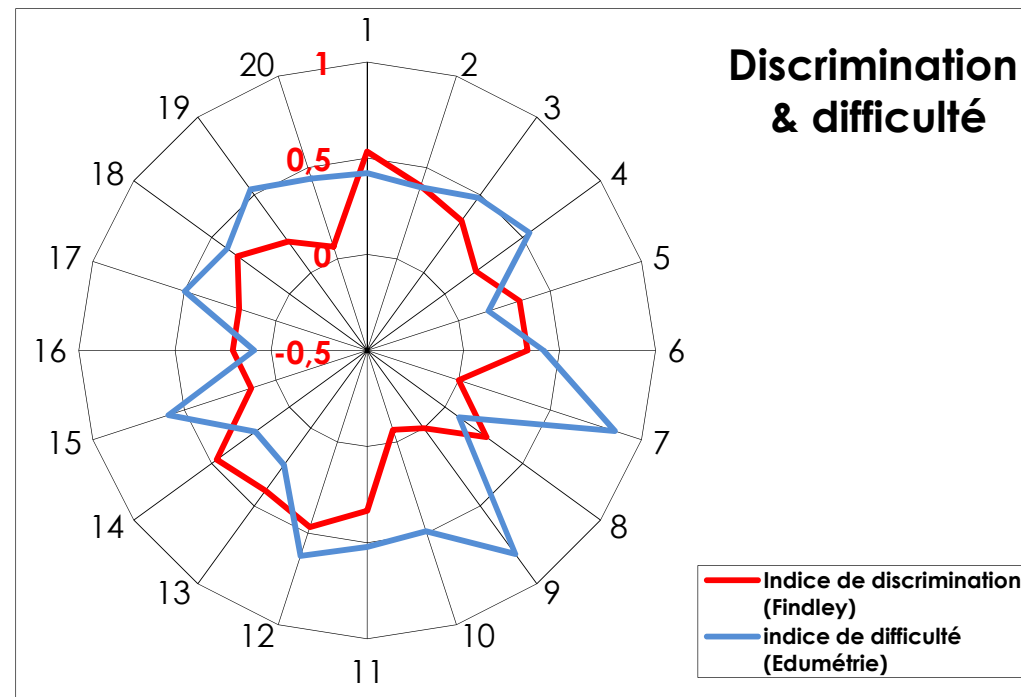


Analyse notes d'écrit 2008

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	10	8	10	9	10	8	5	5	5	6	6	10	10	10	8	6	8	8	10	5	157	31,4
2	8	8	8	9	10	7	0	10	5	7	6	10	10	10	7	10	6	8	4	9	152	30,4
3	6	6	8	3	10	6	5	10	5	6	8	10	10	10	8	10	6	8	8	5	148	29,6
4	10	10	2	8	10	7	5	10	0	5	8	10	10	10	4	10	8	6	4	5	142	28,4
5	8	6	8	3	10	7	5	10	0	5	2	10	10	10	6	10	6	6	10	5	137	27,4
6	10	10	3	7	10	8	5	10	5	7	6	3	0	10	7	10	4	8	8	5	136	27,2
7	6	8	8	8	10	6	5	10	5	5	4	3	10	10	5	10	6	6	0	9	134	26,8
8	8	6	4	5	10	7	5	10	0	6	8	10	10	10	5	6	10	4	4	5	133	26,6
9	8	10	8	1	10	6	0	10	5	6	5	10	10	10	4	10	6	4	4	5	132	26,4
10	6	10	8	8	0	7	5	5	5	7	8	0	10	10	5	10	6	8	4	9	131	26,2
11	10	8	3	7	10	5	0	10	10	5	3	0	10	10	2	10	6	8	8	5	130	26
12	6	8	7	7	10	4	5	10	0	7	8	3	10	10	2	10	6	6	6	5	130	26
13	10	8	6	6	10	8	0	10	0	7	4	3	10	10	6	10	2	8	6	5	129	25,8
14	8	10	5	5	10	8	5	10	0	7	3	0	10	10	4	10	4	8	2	9	128	25,6
15	6	6	6	4	10	8	0	10	0	5	8	10	10	10	6	10	4	4	4	5	126	25,2
16	6	8	8	5	10	8	0	10	0	3	8	0	10	10	5	10	6	8	4	5	124	24,8
17	6	0	8	7	10	7	0	10	5	6	3	0	10	10	7	10	4	6	8	5	122	24,4
18	4	8	8	3	10	5	5	10	0	3	6	3	10	10	4	10	8	8	2	5	122	24,4
19	4	10	6	0	10	6	0	10	5	4	3	6	10	10	4	10	4	6	4	9	121	24,2
20	8	10	10	2	10	4	0	10	0	3	8	10	10	10	4	3	2	4	2	9	119	23,8
21	8	10	4	7	10	6	0	10	0	6	4	3	10	10	1	10	8	4	2	5	118	23,6
22	2	8	10	5	10	4	0	10	0	6	4	6	10	10	5	10	2	6	4	5	117	23,4
23	8	10	4	3	10	7	0	10	0	4	5	0	10	10	6	10	6	6	2	5	116	23,2
24	8	0	6	1	10	7	5	10	0	4	5	3	10	10	8	10	8	6	4	1	116	23,2
25	8	6	7	0	10	6	0	10	10	6	6	10	10	0	0	10	6	6	0	5	116	23,2
26	6	6	6	6	10	7	0	10	0	4	8	3	0	10	7	10	6	8	4	5	116	23,2
27	0	6	8	0	10	5	5	10	0	4	3	3	10	10	3	10	6	8	4	9	114	22,8
28	6	4	3	2	10	4	0	10	5	2	4	10	10	10	7	10	0	4	4	9	114	22,8
29	6	0	7	6	10	6	0	10	5	7	8	6	0	0	5	10	8	8	6	5	113	22,6
30	4	8	8	5	0	7	5	5	0	6	6	3	10	0	3	10	6	8	6	9	109	21,8

31	10	8	2	7	10	7	0	10	0	6	1	3	0	0	5	10	4	6	10	9	108	21,6
32	4	6	2	3	10	6	0	10	0	3	8	3	10	10	0	10	6	8	6	0	105	21
33	8	8	7	9	0	6	0	10	0	4	4	6	10	10	3	3	2	4	6	4	104	20,8
34	4	8	2	8	10	5	0	10	0	4	3	0	10	10	6	10	6	4	2	1	103	20,6
35	4	0	8	2	10	3	0	10	0	6	3	0	10	10	3	10	6	6	6	5	102	20,4
36	6	0	1	8	10	5	0	10	0	5	4	6	0	10	4	10	8	8	2	5	102	20,4
37	4	8	2	4	10	2	0	10	0	6	6	3	10	10	3	10	4	2	4	4	102	20,4
38	6	0	2	8	10	4	0	10	0	4	2	0	10	10	0	10	2	6	6	9	99	19,8
39	2	0	4	0	10	9	0	5	0	5	8	3	10	10	4	10	2	6	6	5	99	19,8
40	6	6	8	3	10	7	0	10	5	7	2	0	0	0	7	10	2	6	4	5	98	19,6
41	4	4	8	3	10	6	0	10	0	3	5	0	0	10	4	10	4	6	6	5	98	19,6
42	8	10	1	7	10	6	0	5	0	6	8	0	10	0	1	6	6	4	8	2	98	19,6
43	6	6	4	5	0	6	0	5	0	6	2	3	10	10	6	10	4	6	2	7	98	19,6
44	6	10	2	4	0	6	0	10	0	2	2	0	10	10	2	10	4	8	6	5	97	19,4
45	8	8	4	8	0	7	0	10	0	3	4	10	0	0	2	10	6	6	4	5	95	19
46	2	10	0	2	10	5	0	5	0	4	3	0	10	10	3	10	4	8	6	1	93	18,6
47	4	0	1	1	10	4	0	10	0	3	6	0	10	10	7	10	4	4	4	2	90	18
48	2	0	5	4	10	6	5	10	10	4	4	3	0	0	1	10	4	6	0	5	89	17,8
49	6	0	5	6	10	2	0	10	0	5	3	3	0	10	4	6	4	8	2	5	89	17,8
50	2	6	8	4	0	4	5	5	5	5	1	0	10	10	2	6	4	4	2	5	88	17,6
51	4	6	4	9	10	4	0	10	0	4	2	0	0	0	3	10	0	6	6	9	87	17,4
52	4	0	0	2	10	6	0	5	5	7	3	0	10	0	4	10	6	0	4	9	85	17
53	0	10	4	1	10	4	0	10	0	6	3	0	0	0	0	6	6	6	8	9	83	16,6
54	4	4	0	1	0	4	0	10	0	8	3	6	10	10	4	0	6	6	2	5	83	16,6
55	0	0	4	0	0	6	0	5	5	6	5	0	0	0	1	10	4	0	6	5	57	11,4

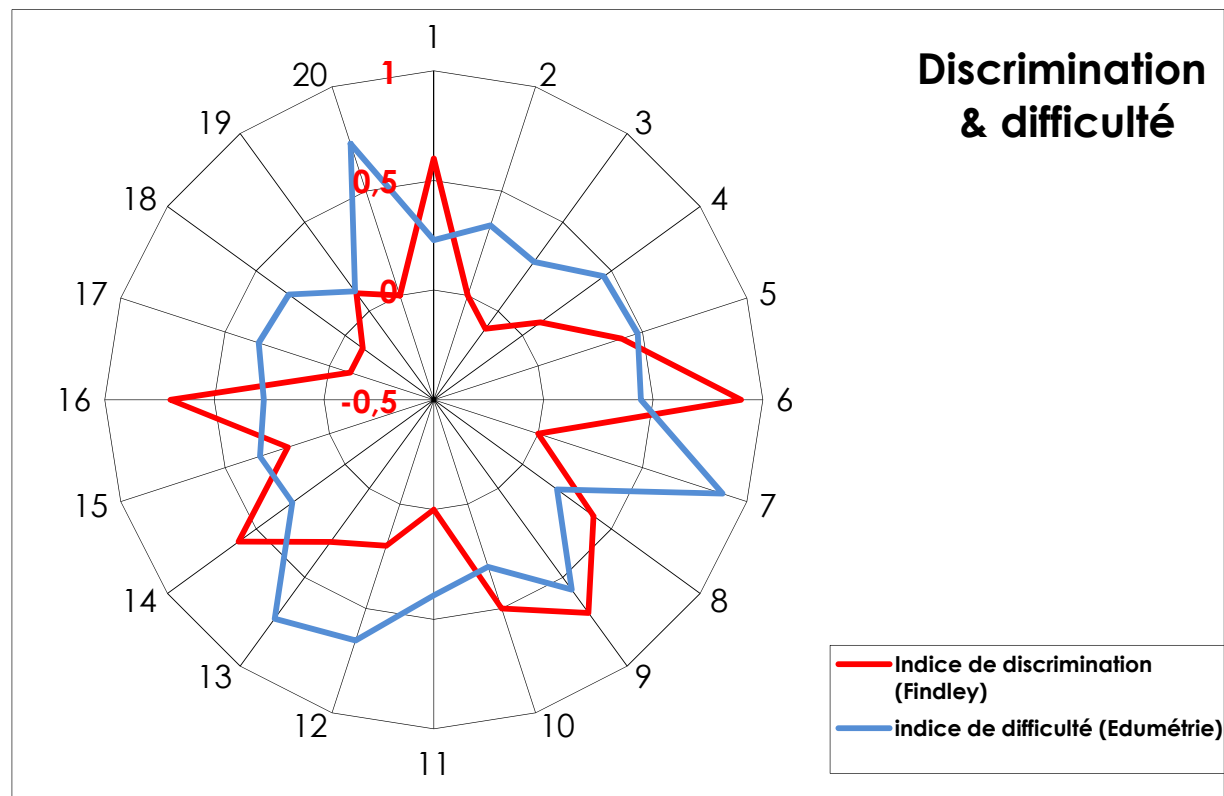
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	5,78	6,07	5,18	4,56	8,36	5,84	1,45	9,09	1,91	5,11	4,78	3,76	7,64	7,82	4,13	9,13	5,02	6	4,65	5,6
Médiane	6	8	5	5	10	6	0	10	0	5	4	3	10	10	4	10	6	6	4	5
MINI	0	0	0	0	0	2	0	5	0	2	1	0	0	0	0	0	0	0	0	0
MAXI	10	10	10	9	10	9	5	10	10	8	8	10	10	10	8	10	10	8	10	9
Réussite des "FORTS" (>8)	0,67	0,73	0,47	0,33	0,93	0,33	0	0,87	0,07	0	0,4	0,53	0,93	1	0,13	0,87	0,2	0,53	0,33	0,27
Réussite des "FAIBLES"(>8)	0,13	0,33	0,13	0,13	0,6	0	0	0,6	0,07	0,07	0,07	0,07	0,53	0,53	0	0,67	0	0,2	0,13	0,2
Indice de discrimination (Findley)	0,53	0,40	0,33	0,20	0,33	0,33	0,00	0,27	0,00	0,07	0,33	0,47	0,40	0,47	0,13	0,20	0,20	0,33	0,20	0,07
indice de difficulté (Edumétrie)	0,42	0,39	0,48	0,54	0,16	0,42	0,85	0,09	0,81	0,49	0,52	0,62	0,24	0,22	0,59	0,09	0,50	0,40	0,53	0,44
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,68	0,37	0,49	0,54	0	0,8	0,39	0,24	0,37	1,05	0,83	0,56	0	0	0,78	0,2	0,93	0,8	0,78	0,83



Analyse notes d'écrit 2009

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	10	8	7	6	8	10	6	10	10	8	5	9	10	10	10	9	7	7	10	6	166	33,2
2	7	10	10	2	8	8	6	10	10	10	6	3	5	8	10	8	7	8	10	5	151	30,2
3	10	0	7	8	8	8	0	10	10	9	8	4	5	10	10	9	8	8	10	5	147	29,4
4	10	10	9	7	6	8	4	10	5	7	7	5	5	8	10	8	7	6	10	3	145	29
5	10	8	9	6	6	10	0	10	10	9	7	3	5	8	8	7	7	8	10	2	143	28,6
6	10	10	7	6	4	10	0	10	10	8	7	4	5	6	10	8	7	5	10	3	140	28
7	10	0	7	8	6	10	0	10	10	6	5	9	5	10	10	7	7	6	10	3	139	27,8
8	10	8	8	6	6	10	0	10	10	6	5	6	10	10	8	7	7	6	0	2	135	27
9	10	8	9	8	4	10	2	10	5	8	9	2	1	8	8	9	4	5	10	5	135	27
10	10	0	5	4	8	10	6	0	10	8	7	2	10	10	0	9	6	7	10	6	128	25,6
11	4	10	10	4	6	10	2	10	5	7	6	5	5	9	0	6	6	6	10	1	122	24,4
12	10	8	5	6	6	10	0	10	0	4	8	4	0	5	10	6	6	8	10	4	120	24
13	4	8	8	6	4	6	2	10	0	6	4	1	10	10	8	6	7	7	10	3	120	24
14	8	10	8	2	8	10	6	0	5	5	5	4	0	7	8	7	7	7	10	2	119	23,8
15	10	8	7	2	6	0	0	10	5	9	8	6	5	8	0	9	7	8	10	0	118	23,6
16	10	8	5	10	4	10	0	10	5	7	6	2	0	9	10	8	6	5	0	3	118	23,6
17	8	9	8	2	2	6	2	10	10	6	6	4	5	2	0	8	7	7	10	4	116	23,2
18	8	0	9	6	4	0	0	10	0	6	6	3	5	10	10	8	7	6	10	6	114	22,8
19	8	8	5	8	8	10	0	10	0	6	6	4	0	7	0	8	7	6	10	3	114	22,8
20	10	0	8	8	2	10	0	10	5	6	6	1	0	8	8	7	4	8	10	3	114	22,8
21	4	8	8	2	6	4	0	10	5	8	5	3	0	8	8	7	8	9	10	0	113	22,6
22	10	10	0	6	8	2	0	10	5	8	7	5	0	4	10	6	7	5	10	0	113	22,6
23	4	10	6	4	10	0	2	10	0	9	6	3	0	6	0	7	8	9	10	5	109	21,8
24	7	0	5	6	6	8	0	10	0	6	7	3	0	6	10	7	7	8	10	1	107	21,4
25	8	8	9	8	2	2	2	10	0	8	4	2	0	2	8	7	7	3	10	5	105	21
26	6	0	9	10	6	0	0	10	5	9	7	2	0	4	0	8	7	9	10	2	104	20,8
27	8	8	5	6	2	8	2	0	10	5	7	3	0	7	0	7	7	7	10	1	103	20,6
28	6	6	4	8	2	2	0	10	5	5	6	5	5	8	10	6	6	8	0	0	102	20,4
29	8	8	8	6	6	2	0	0	0	6	4	4	0	8	10	7	2	8	10	3	100	20
30	6	0	9	2	2	0	0	10	0	8	8	0	0	7	10	6	6	8	10	3	95	19
31	4	6	6	6	6	0	0	10	0	4	4	3	0	7	10	6	5	5	10	3	95	19
32	6	8	8	2	2	0	0	10	0	10	7	2	0	4	0	7	8	6	10	4	94	18,8
33	6	10	7	2	4	0	0	10	0	6	4	4	0	3	8	7	7	6	10	0	94	18,8
34	6	8	8	8	6	0	0	0	0	7	9	1	0	8	0	6	7	7	10	1	92	18,4
35	8	10	9	2	4	0	0	0	0	6	4	3	0	6	8	6	7	6	10	0	89	17,8
36	4	8	8	2	2	6	0	0	0	6	3	1	0	2	10	7	8	8	0	1	76	15,2

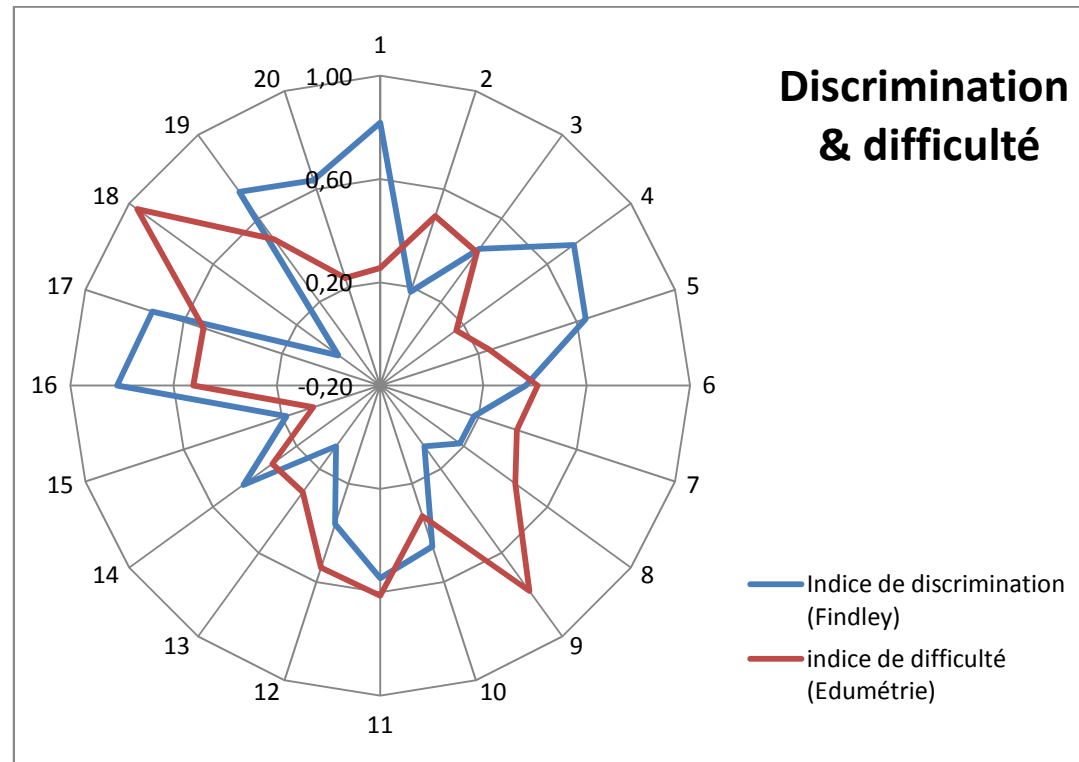
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	7,72	6,64	7,22	5,42	5,22	5,56	1,17	8,06	4,31	7	6,08	3,47	2,67	7,03	6,67	7,25	6,61	6,83	8,89	2,72
Médiane	8	8	8	6	6	7	0	10	5	7	6	3	0	8	8	7	7	7	10	3
MINI	4	0	0	2	2	0	0	0	0	4	3	0	0	2	0	6	2	3	0	0
MAXI	10	10	10	10	10	10	6	10	10	10	9	9	10	10	10	9	8	9	10	6
Réussite des "FORTS"(>8)	0,9	0,7	0,5	0,3	0,4	1	0	0,9	0,8	0,7	0,2	0,2	0,3	0,9	0,9	0,7	0,1	0,3	0,9	0
Réussite des "FAIBLES"(>8)	0,3	0,7	0,6	0,2	0	0,1	0	0,5	0,1	0,2	0,2	0	0	0,3	0,7	0	0,2	0,4	0,8	0
Indice de discrimination (Findley)	0,60	0,00	0,10	0,10	0,40	0,90	0,00	0,40	0,70	0,50	0,00	0,20	0,30	0,60	0,20	0,70	0,10	0,10	0,10	0,00
indice de difficulté (Edumétrie)	0,23	0,34	0,28	0,46	0,48	0,44	0,88	0,19	0,57	0,30	0,39	0,65	0,73	0,30	0,33	0,28	0,34	0,32	0,11	0,73
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,33	0,06	0,25	0,42	0,56	0,11	0,14	0	0,31	0,47	0,58	0,64	0,31	0,25	0	0,25	0,25	0,42	0	0,58



Analyse notes d'écrit 2010

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	10	6	8	10	8	10	8	6	7	10	10	8	0	9	10	10	10	4	10	10	164	32,8
2	8	7	8	10	10	10	8	7	6	7	0	7	10	8	10	10	10	4	10	10	160	32
3	8	7	8	10	10	10	8	6	2	7	4	10	10	8	10	10	10	0	10	10	158	31,6
4	10	7	8	10	8	8	8	5	8	8	6	10	10	7	10	10	0	0	10	10	153	30,6
5	10	10	8	10	10	8	8	7	2	9	10	6	0	7	10	10	10	0	10	6	151	30,2
6	9	5	6	10	10	8	6	7	3	8	6	9	0	9	10	10	10	4	10	10	150	30
7	9	7	6	10	10	6	8	6	2	10	10	4	10	7	10	10	4	0	10	10	149	29,8
8	10	7	8	8	0	10	6	8	2	6	10	5	10	8	10	10	10	0	10	10	148	29,6
9	8	10	6	8	10	6	8	8	2	6	10	6	10	9	10	10	10	0	0	8	145	29
10	10	6	5	10	10	10	6	6	3	5	6	5	10	9	10	0	10	0	10	10	141	28,2
11	7	0	5	10	10	6	6	4	2	8	10	7	10	6	10	10	10	0	10	10	141	28,2
12	8	7	4	10	6	0	8	4	5	6	10	4	10	7	10	10	10	0	10	8	137	27,4
13	8	7	4	10	10	8	6	8	0	8	10	4	10	7	10	0	10	0	10	6	136	27,2
14	8	7	8	10	10	10	4	6	0	7	6	3	10	5	10	10	10	0	0	10	134	26,8
15	10	7	5	10	10	2	6	4	0	8	6	2	10	8	10	10	0	0	10	10	128	25,6
16	5	7	8	10	8	6	6	5	2	6	0	5	10	7	10	0	10	4	10	6	125	25
17	9	3	3	8	4	8	10	7	3	6	0	2	10	10	10	0	10	0	10	10	123	24,6
18	9	6	5	10	10	10	10	4	2	6	6	8	10	7	10	0	0	0	0	10	123	24,6
19	9	10	9	10	10	6	4	6	1	5	6	8	10	7	10	0	0	0	0	10	121	24,2
20	10	7	6	10	8	8	4	5	7	8	0	6	10	8	10	0	4	0	0	10	121	24,2
21	6	3	7	10	10	8	4	5	6	8	10	5	0	8	10	0	0	0	10	10	120	24
22	8	6	4	6	8	6	6	6	0	5	10	5	0	9	6	0	10	0	10	10	115	23
23	8	5	8	10	10	4	2	6	3	5	0	6	10	7	10	10	4	0	0	6	114	22,8
24	5	0	7	10	2	4	6	8	2	7	0	2	10	8	10	10	10	0	0	10	111	22,2
25	3	3	5	10	10	8	8	5	1	5	0	4	0	9	10	10	0	0	10	10	111	22,2
26	8	7	4	6	10	2	6	5	3	10	0	9	0	9	10	10	0	0	0	10	109	21,8
27	7	7	3	7	10	6	6	3	2	4	10	5	10	7	10	10	0	0	0	0	107	21,4
28	7	7	4	6	10	2	6	3	0	7	6	0	0	3	10	10	4	0	10	10	105	21
29	8	10	5	7	10	2	6	4	6	7	0	6	10	8	6	0	0	0	0	6	101	20,2
30	8	7	8	10	10	2	4	8	3	8	0	6	0	2	10	0	4	0	0	6	96	19,2
31	8	10	5	5	8	4	10	7	0	6	0	0	10	4	8	0	0	0	0	10	95	19
32	7	0	8	10	0	2	8	5	2	8	0	4	10	7	10	0	0	0	10	0	91	18,2
33	8	0	3	8	8	8	2	4	0	4	0	2	10	7	10	0	4	0	0	10	88	17,6
34	4	7	4	6	6	8	8	3	0	7	0	0	0	2	10	0	0	0	10	10	85	17
35	7	0	3	5	4	2	8	4	0	7	0	3	10	5	8	0	10	0	0	6	82	16,4
36	3	1	4	4	2	8	10	6	0	4	0	4	10	6	10	0	4	0	0	6	82	16,4
37	5	2	6	7	2	6	6	7	1	6	0	1	10	2	10	10	0	0	0	0	81	16,2
38	7	0	6	10	6	0	4	4	1	7	0	5	0	8	10	0	0	0	0	10	78	15,6
39	6	0	6	7	10	0	4	2	1	7	0	3	10	8	8	0	0	0	0	0	72	14,4
40	3	0	5	5	0	6	6	7	0	5	0	2	10	5	8	0	0	0	0	0	62	12,4
41	6	0	1	3	8	2	8	6	1	5	0	0	0	4	6	0	4	0	0	6	60	12
42	6	6	2	5	0	8	4	5	0	5	0	2	0	6	0	0	0	0	0	0	49	9,8

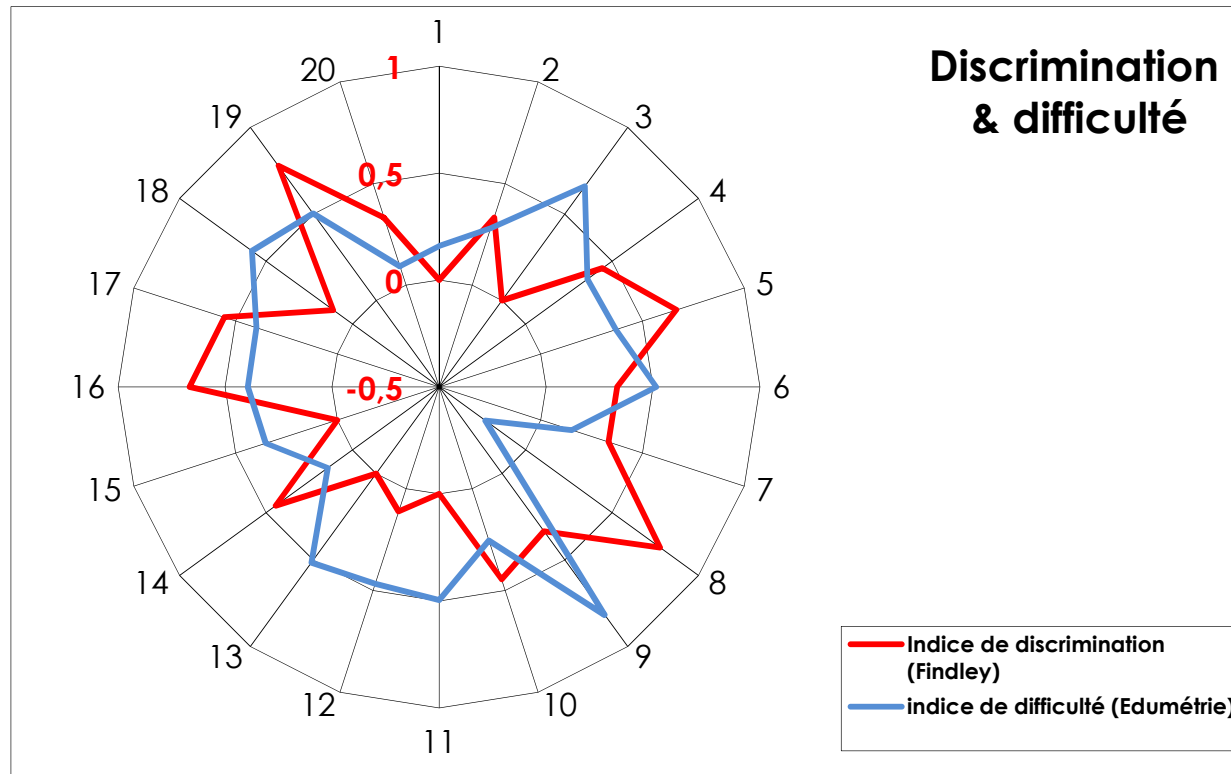
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	7,45	5,1	5,62	8,36	7,52	5,9	6,43	5,52	2,17	6,69	3,86	4,6	6,9	6,83	9,29	4,76	4,81	0,38	5	7,62
Médiane	8	6,5	5,5	10	10	6	6	6	2	7	0	5	10	7	10	0	4	0	5	10
MINI	3	0	1	3	0	0	2	2	0	4	0	0	0	2	0	0	0	0	0	0
MAXI	10	10	9	10	10	10	10	8	8	10	10	10	10	10	10	10	10	4	10	10
Réussite des "FORTS" (>8)	0,91	0,18	0,55	1	0,91	0,73	0,64	0,18	0,09	0,55	0,55	0,36	0,73	0,64	1	0,91	0,82	0	0,91	0,91
Réussite des "FAIBLES"(>8)	0,09	0	0,09	0,27	0,27	0,36	0,45	0	0	0,09	0	0	0,64	0,18	0,82	0,09	0,09	0	0,18	0,27
Indice de discrimination (Findley)	0,82	0,18	0,45	0,73	0,64	0,36	0,18	0,18	0,09	0,45	0,55	0,36	0,09	0,45	0,18	0,82	0,73	0,00	0,73	0,64
indice de difficulté (Edumétrie)	0,25	0,49	0,44	0,16	0,25	0,41	0,36	0,45	0,78	0,33	0,61	0,54	0,31	0,32	0,07	0,52	0,52	0,96	0,50	0,24
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,26	0,24	0,63	0,24	0,12	0,29	0,56	0,71	0,24	0,46	0,22	0,54	0	0,22	0,07	0	0,2	0,1	0	0,22



Analyse notes d'écrit 2011

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Total	Note/40
1	10	8	4	10	10	10	10	16	0	10	6	8	7	10	7	10	10	6	10	10	172	34,4
2	9	6	4	10	6	10	10	16	10	8	6	5	6	10	7	10	7	6	10	10	166	33,2
3	9	8	4	10	10	10	10	16	0	8	4	5	5	10	8	10	9	8	10	10	164	32,8
4	9,5	9	4	10	10	0	10	16	0	10	6	7	7	10	5	10	7	2	10	10	153	30,5
5	10	6	4	0	7	10	10	16	10	6	4	4	6	8	6	10	9	8	6	10	150	30
6	9	10	0	10	6	0	10	16	10	8	4	5	6	10	6	10	4	4	10	10	148	29,6
7	10	9	4	10	10	10	10	16	0	8	4	6	5	6	9	0	10	6	4	10	147	29,4
8	5,5	7	4	6	8	10	8	16	10	8	4	7	4	10	2	0	10	6	10	10	146	29,1
9	10	6	4	10	10	0	10	16	0	8	4	5	3	10	6	10	9	4	10	10	145	29
10	7,5	9	0	10	6	10	6	16	10	8	6	4	4	5	7	10	6	4	2	10	141	28,1
11	9	6	4	10	10	10	9	16	0	8	6	5	4	10	8	10	3	0	2	10	140	28
12	9,5	9	4	10	7	0	10	16	0	8	7	8	5	10	7	0	9	4	8	8	140	27,9
13	9	9	4	10	5	0	10	16	0	10	5	5	4	10	5	10	6	4	6	10	138	27,6
14	8,5	7	4	10	7	10	7	16	0	8	6	5	6	10	6	10	5	0	2	10	138	27,5
15	9	7	4	10	8	0	10	16	0	6	7	3	4	7	7	10	2	8	6	10	134	26,8
16	8,5	5	4	10	4	10	9	16	0	6	4	4	7	10	6	10	10	0	0	10	134	26,7
17	7	6	4	10	6	0	10	16	0	6	5	4	3	10	10	10	3	4	10	8	132	26,4
18	6	6	4	0	5	10	9	16	0	10	4	6	7	6	7	10	3	6	6	10	131	26,2
19	6	6	4	10	5	0	9	16	0	8	5	10	5	5	6	10	9	2	4	8	128	25,6
20	10	10	4	4	8	0	10	16	0	8	5	7	3	7	6	0	7	2	10	10	127	25,4
21	6	9	4	0	8	10	9	16	0	10	6	3	4	10	5	10	2	6	2	7	127	25,4
22	10	7	4	0	6	10	10	0	0	8	6	6	6	10	8	10	9	4	2	10	126	25,2
23	8,5	8	0	0	7	0	4	16	0	8	6	8	7	10	7	10	7	8	0	10	125	24,9
24	7,5	8	4	6	8	0	10	16	0	6	4	7	5	10	7	0	2	2	10	10	123	24,5
25	9,5	7	4	10	3	0	10	0	0	10	5	4	5	10	6	10	9	0	2	10	115	22,9
26	8	5	4	0	4	0	8	16	0	10	7	6	4	10	4	0	5	6	6	10	113	22,6
27	7,5	8	4	0	6	10	10	16	0	4	5	4	4	6	6	0	6	4	2	10	113	22,5
28	8,5	3	0	10	4	0	10	0	10	8	4	3	4	10	8	0	6	4	2	10	105	20,9
29	7,5	8	4	10	3	0	10	7	0	2	2	4	5	10	6	0	7	2	0	9	96,5	19,3
30	8	6	4	0	4	10	10	0	0	6	3	5	6	6	7	0	5	6	2	8	96	19,2
31	9,5	7	4	0	5	10	1	0	0	8	7	6	5	7	7	0	3	8	0	5	92,5	18,5
32	7	7	4	6	2	0	1	0	0	0	3	4	2	6	4	0	0	2	0	6	54	10,8
33	8	6	0	0	0	0	0	0	0	6	6	3	2	4	8	0	0	2	0	0	45	9

N°	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
MOYENNE / 10	8,41	7,21	3,39	6,42	6,3	4,85	8,48	12,3	1,82	7,45	5,03	5,33	4,85	8,58	6,48	6,06	6,03	4,18	4,97	9,06
Médiane	8,5	7	4	10	6	0	10	16	0	8	5	5	5	10	7	10	6	4	4	10
MINI	5,5	3	0	0	0	0	0	0	0	0	2	3	2	4	2	0	0	0	0	0
MAXI	10	10	4	10	10	10	10	16	10	10	7	10	7	10	10	10	10	8	10	10
Réussite des "FORTS"(>8)	0,89	0,56	0	0,78	0,67	0,67	1	1	0,44	0,89	0	0,11	0	0,89	0,22	0,78	0,67	0,22	0,78	1
Réussite des "FAIBLES"(>8)	0,89	0,22	0	0,33	0	0,33	0,67	0,22	0,11	0,44	0	0	0	0,44	0,22	0,11	0,11	0,11	0	0,67
Indice de discrimination (Findley)	0,00	0,33	0,00	0,44	0,67	0,33	0,33	0,78	0,33	0,44	0,00	0,11	0,00	0,44	0,00	0,67	0,56	0,11	0,78	0,33
indice de difficulté (Edumétrie)	0,16	0,28	0,66	0,36	0,37	0,52	0,15	-0,23	0,82	0,25	0,50	0,47	0,52	0,14	0,35	0,39	0,40	0,58	0,50	0,09
"Pas" de notation																				
Résultats "MOYENS" (de 3 à 7)	0,12	0,36	0,85	0,12	0,48	0	0,06	0	0	0,24	0,85	0,76	0,79	0,24	0,45	0	0,36	0,52	0,21	0,06



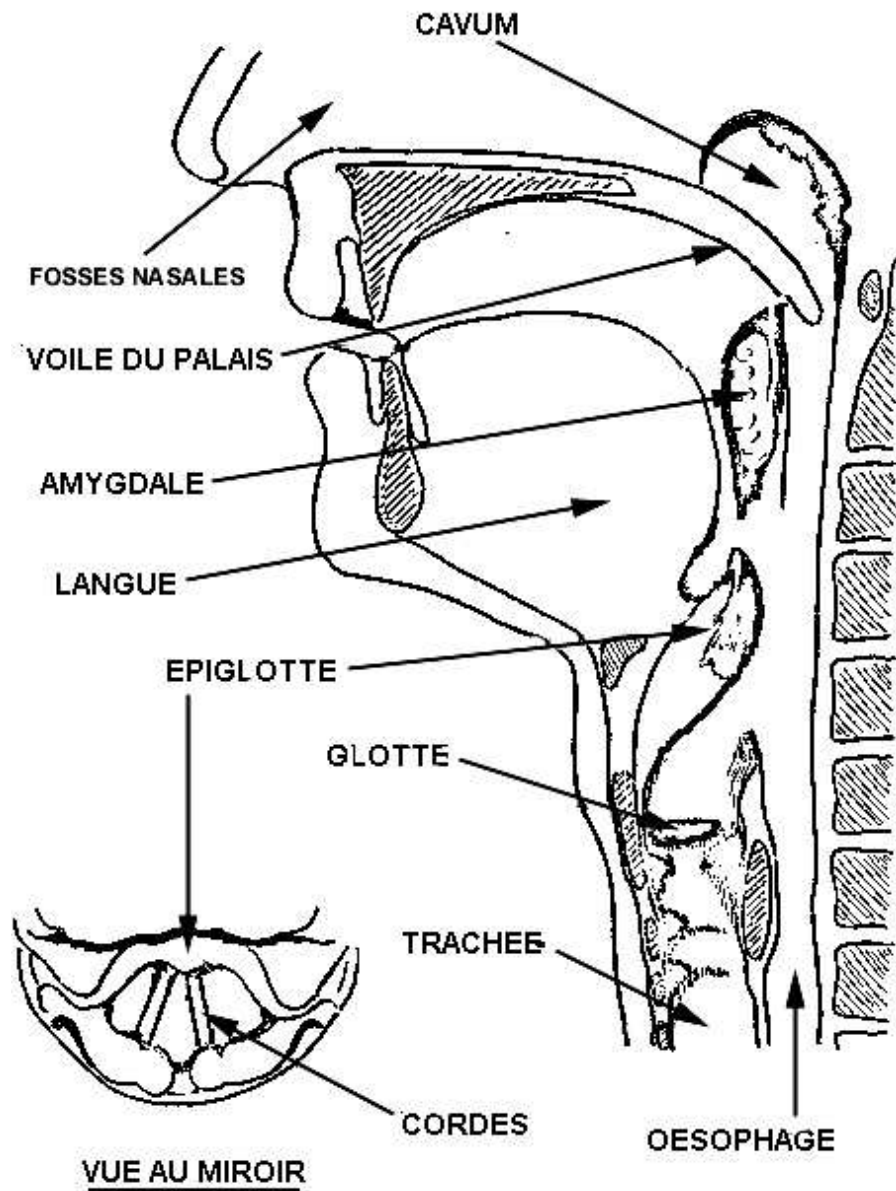
VI. Annexe 6

22 Questions Discriminantes :

2007

7-2 Légendez le schéma suivant en indiquant vos réponses dans les dix cadres prédéfinis ci-dessous.

...../ 10 pts



7-2 Pour répondre aux 10 affirmations du tableau ci-dessous, mettez une croix dans les cases correspondant aux items VRAI ou FAUX.

...../ 10 pts

AFFIRMATIONS	VRAI	FAUX
La calcémie n'est pas élevée dans l'insuffisance parathyroïdienne.	X	
Le temps de coagulation correspond au temps de formation du clou plaquettaire.		X
La kaliémie est abaissée au cours des états de chocs.		X
Au cours des diarrhées prolongées, la natrémie peut être élevée.	X	
La vitesse de sédimentation indique la vitesse de croissance des globules blancs.		X
La réserve alcaline est augmentée dans l'acidose métabolique.		X
Chez le nouveau-né, la crise hémolytique favorise l'augmentation du nombre des hématies.		X
Le neurotransmetteur des jonctions neuromusculaires des muscles striés est la succinylcholine.		X
Toute diminution de la PaO_2 inférieure à 60 mmHg entraîne une diminution de la ventilation.		X
Chez l'adulte, la pression normale du liquide céphalo-rachidien est comprise entre 0 et 5 cm H_2O .	X	

7-4 La loi n° 2002-303 du 4 mars 2002 « relative aux droits des malades et à la qualité du système de santé », donne aux patients un certain nombre de droits : citez-en deux.

...../ 10 pts

- l'accès au dossier médical.
- les nouveaux droits des mineurs concernant leur secret médical vis à vis de l'autorité parentale.
- renforcement de la place des associations de patients et usagers dans la gestion du système de santé.
- suspension des peines pour les condamnés très malades.
- l'indemnisation des accidents médicaux.

7-8- Complétez le tableau suivant :

...../ 10 pts

Paramètre sanguin	Valeur normale chez l'homme
Globules rouges	4,2 à 4,7 million /mm ³
Globules blancs	4000 à 10000 / mm ³
Hématocrite	40 à 52 %
Hémoglobine	13 à 18 g /100ml
Plaquettes	150000 à 350000 / mm ³
% de prothrombine	80 à 100 %
Créatinine	7 à 13 mg/l ou 40 à 120 µmol/l
Sodium	138 à 142 mmol/ ou mEq/l
Potassium	3,5 à 4,9 mmol/l ou mEq/l
Bicarbonates	26 28 mmol/l ou mEq/l

7-9 La surveillance d'un traitement par héparine de bas poids moléculaire s'effectue par :

...../ 10 pts

- Temps de céphaline activé ☐
- Numération plaquettaire ☒
- INR ☐
- Dosage de l'anti Xa ☒

(Faites une croix dans la ou les cases correspondantes)

7-10 Vous êtes infirmier(ère) en salle de surveillance post-interventionnelle. Nous sommes le 1^{er} janvier, il est 13 heures. Le médecin prescrit une perfusion continue de NAROPENE® à 2⁰/₀₀ sur le cathéter d'anesthésie loco-régionale d'un patient de 80 kg, à raison de 100µg/kg/h. Le médicament est conditionné dans une poche de 100 ml.

...../ 10 pts

c) Quel sera le débit de perfusion ?



















4 ml/h

d) A quelle date et à quelle heure pouvez-vous prévoir la fin de cette perfusion ?

Le 2 janvier à 14 heures


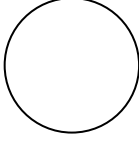
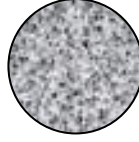

7-13 En examinant les cross-matches suivants, parmi les six possibilités ci-dessous, indiquez celle qui correspond au groupe A et celle qui correspond au groupe O, en notant A ou O dans les cases de la colonne « groupe ».

...../ 10 pts

GROUPE	Sérum anti-A	Sérum anti-B	Sérum anti-AB
			
			
A			
			
			
O			

7-14 Lors du contrôle pré-transfusionnel ultime au lit d'un patient, vous obtenez les résultats suivants :

...../ 10 pts

	Sang du patient	Sang du culot
Sérum ANTI-A		
Sérum ANTI-B		

a) Quel est le groupe sanguin du patient ?

AB

b) Quel est le groupe sanguin du culot ?

B

c) La transfusion est-elle possible ?

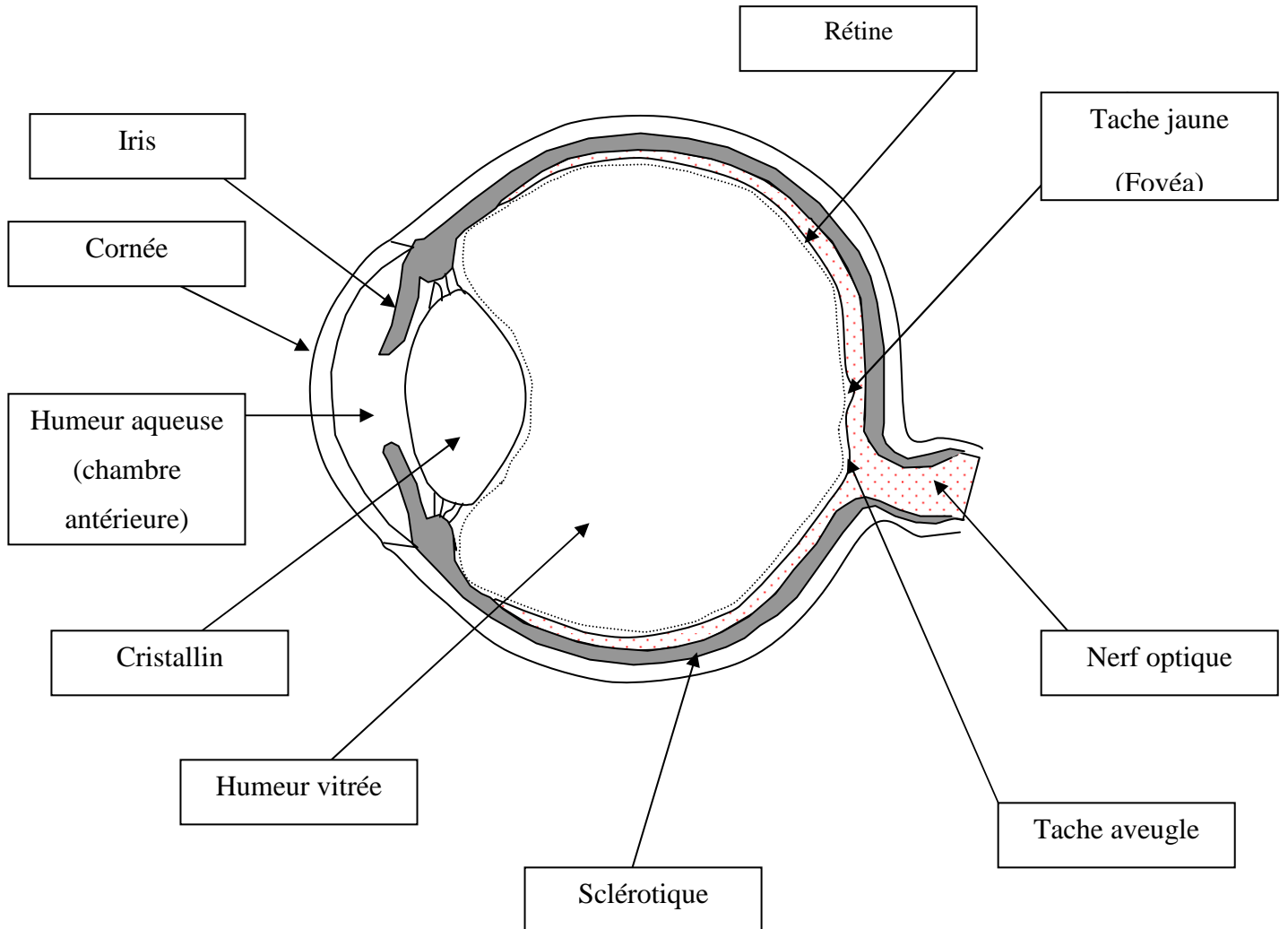
Oui

d) Pourquoi ?

Le culot (groupe B) ne possède pas d'antigène risquant d'agglutiner avec le sang du patient (groupe AB)

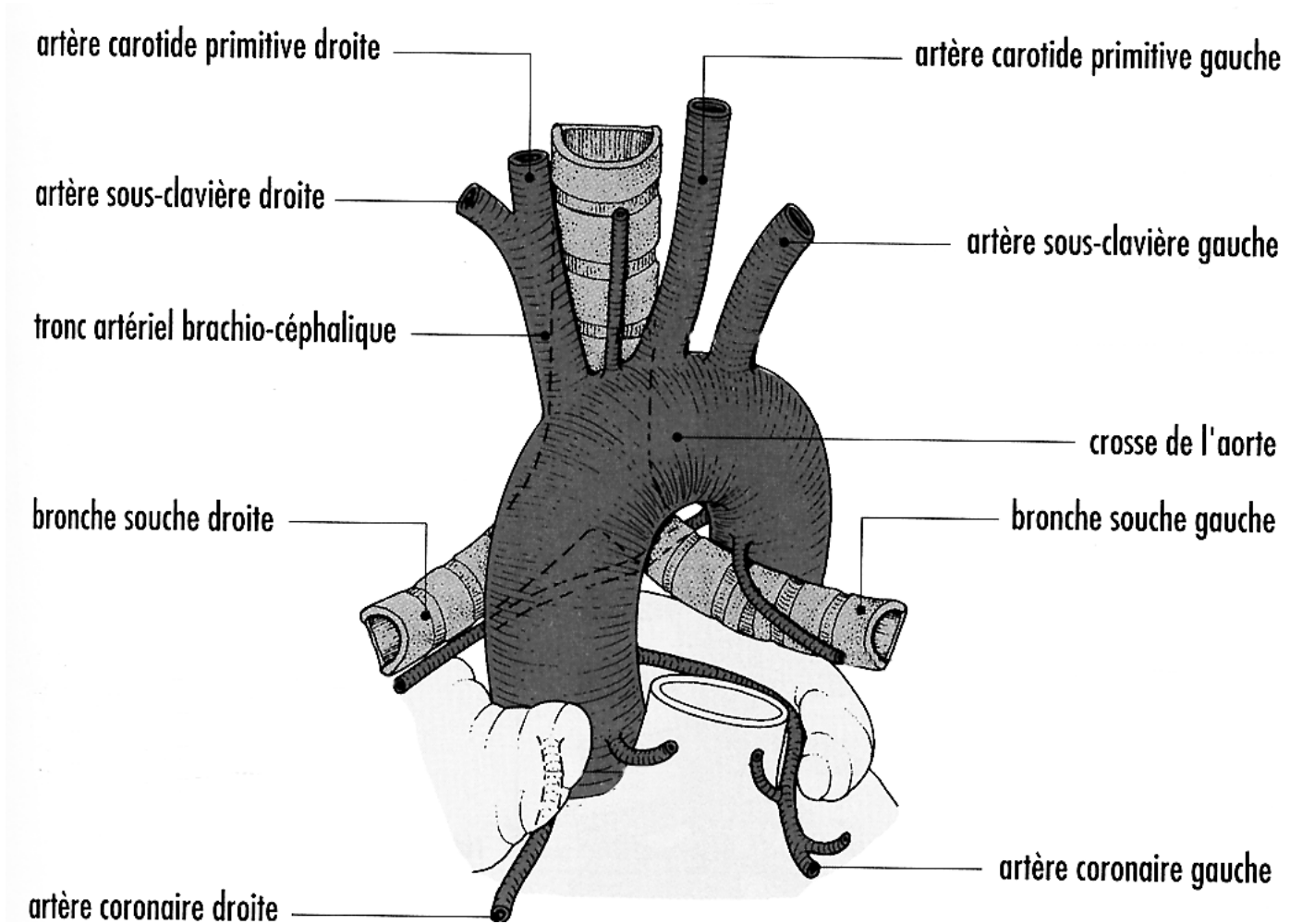
7-17 Légendez cette coupe sagittale de l'œil dans les dix cadres prédéfinis ci-dessous.

...../ 10 pts

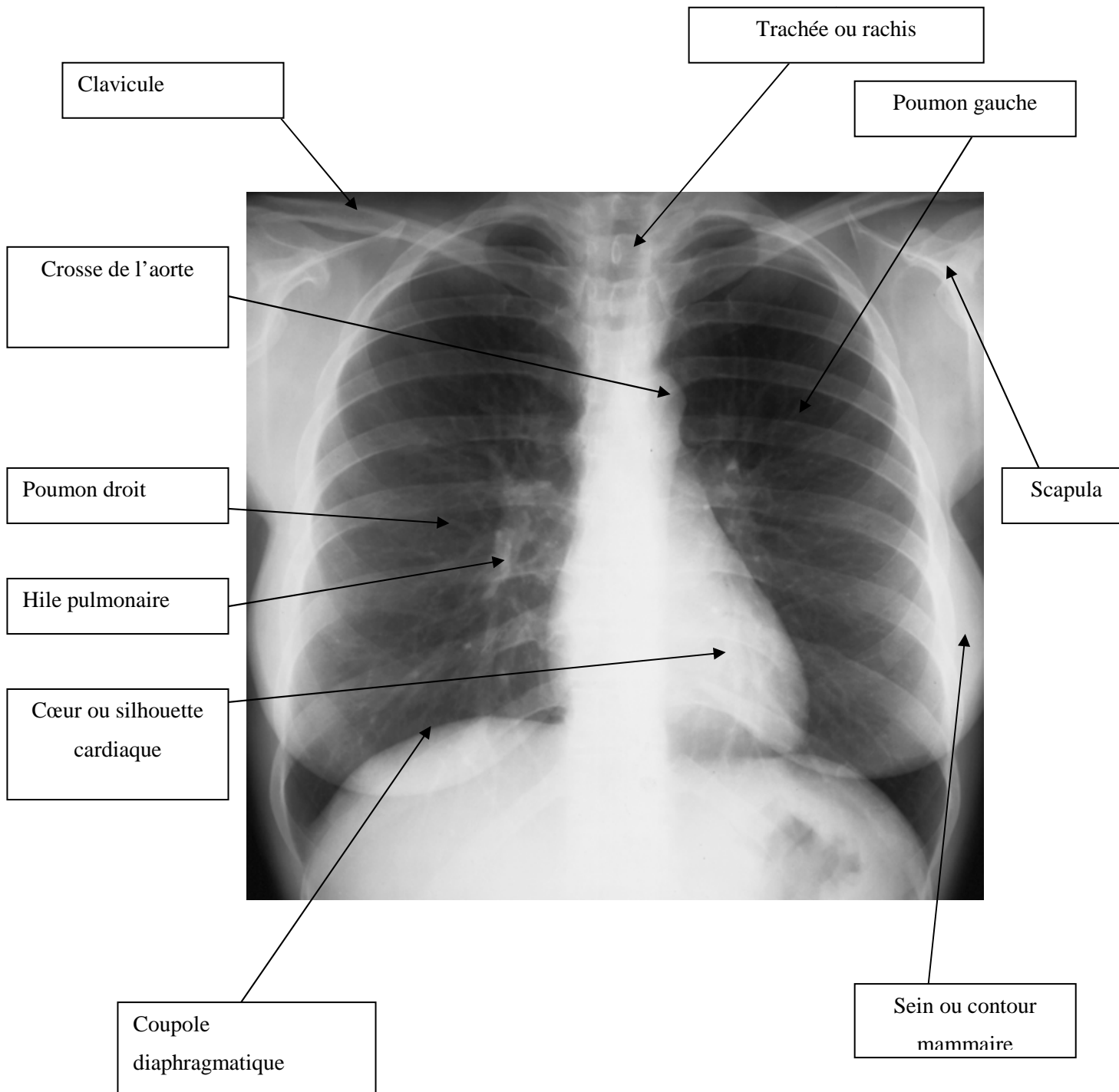


7-18 Légendez le schéma suivant en indiquant vos réponses dans les dix cadres prédéfinis ci-dessous.

...../ 10 pts



7-20 Légendez cette radiographie pulmonaire de face dans les dix cadres prédéfinis.
...../ 10 pts



2010

10-1 Cochez la case correspondant à la bonne réponse.

...../ 10 pts

k- La formation de glucose à partir d'acide aminés est :

- La glycogénolyse ☐
- La néoglucogénèse ☒X
- La glycogénèse ☐

l- Le volume de sang contenu dans les veines est de l'ordre de :

- 40% du sang de l'organisme ☐
- 75% du sang de l'organisme ☐
- 60% du sang de l'organisme ☒X

m- L'air contient :

- 51% d'oxygène ☐
- 21% d'oxygène ☒X
- 25% d'oxygène ☐

n- La stimulation du nerf vague provoque :

- Une augmentation de la pression artérielle ☐
- Une augmentation de la fréquence cardiaque ☐
- Une diminution de la fréquence cardiaque ☒X

o- La concentration en pourcentage d'une solution contenant 0,08g/ml est de :

- 0,08% ☐
- 0,8% ☐
- 8% ☒X

p- Dans le foie la rénine transforme :

- L'angiotensine I en angiotensine II ☐
- L'angiotensinogène en angiotensine I ☒X
- L'angiotensine II en aldostérone ☐

q- Le L.C.R est contenu au niveau de :

- L'espace sous arachnoïdien ☒X
- L'espace sous dural ☐
- L'espace périural ☐

r- La C.R.F (capacité résiduelle fonctionnelle) est égale à:

- volume résiduel + volume de réserve inspiratoire ☐
- volume courant + volume de réserve expiratoire ☐
- volume résiduel + volume de réserve expiratoire ☒X

s- le nombre de paires racines nerveuses d'origine rachidienne est :

- 12 ☐
- 25 ☐
- 31 ☒X

t- Le test d'exploration de l'hémostase primaire est :

- Le temps de saignement ☒X
- Le temps de céphaline + activateur ☐
- Le dosage de fibrinogène ☐

10-3 Complétez le tableau suivant en vous inspirant de l'exemple noté sur la première ligne :

...../ 10 pts

Définitions	Noms
EXEMPLE : <i>Fréquence cardiaque élevée</i>	Tachycardie
Sécrétion d'urine en quantité abondante	Polyurie
Mictions très fréquentes	Pollakiurie
Infection du rein avec rétention de pus dans le bassinet distendu	Pyonéphrose
Abouchement des uretères à la peau	Urétérostomie
Abouchement des cavités pyélocalicielles à la peau	Néphrostomie
Lésion inflammatoire de l'épididyme et du testicule	Orchi-épididymite
Résection totale ou partielle de la vessie	Cystectomie
Incapacité du rein à assurer ses fonctions avec accumulation d'urée, de créatinine et souvent de potassium	Insuffisance rénale
Difficultés à la miction	Dysurie
Présence de sang dans les urines	Hématurie

10-4 A partir des deux listes ci-dessous, complétez le tableau de manière à attribuer à chaque médicament sa classe pharmacologique
...../ 10 pts

	Médicament (DCI)
1	Nalbuphine
2	Atropine
3	Vérapamil
4	Salbutamol
5	Furosémide
6	Ropivacaïne
7	Dextran
8	Propanolol
9	Kétamine
10	Midazolam

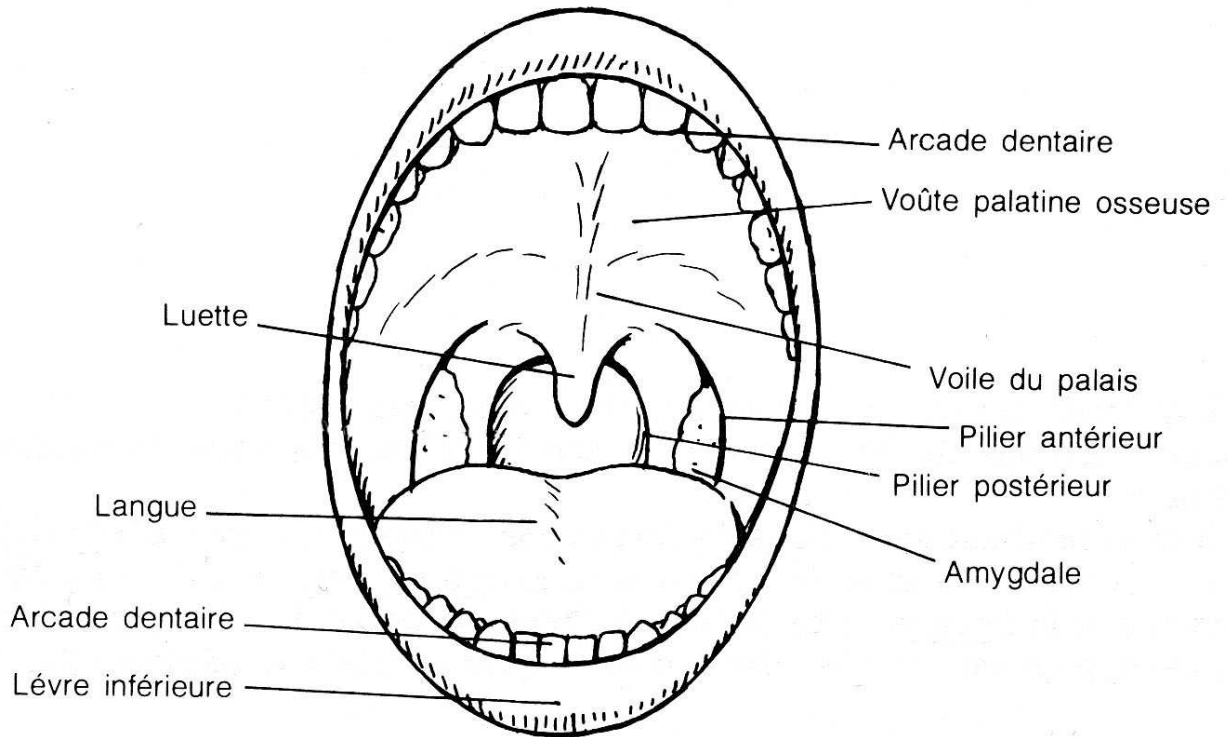
	Classe pharmacologique
A	Benzodiazépine
B	β_2 - mimétique
C	Anesthésique local
D	Anesthésique général
E	Béta-bloquant
F	Agoniste-antagoniste morphinique
G	diurétique
H	Solution colloïdale
I	Parasympathomimétique
J	Inhibiteur calcique

1	2	3	4	5	6	7	8	9	10
F	I	J	B	G	C	H	E	D	A

10-5 Parmi les médicaments suivants, lequel ou lesquels peuvent induire une mydriase ou un myosis ? (Complétez le tableau suivant par des croix dans les cases correspondantes)
...../ 10 pts

Médicaments	Mydriase	Myosis
Morphine		X
Adrénaline	X	
Phényléphrine (Néosynéphrine®)	X	
Atropine	X	
Buprénorphine (Temgésic®)		X

10-10 Légendez le schéma ci-dessous
...../ 10 pts



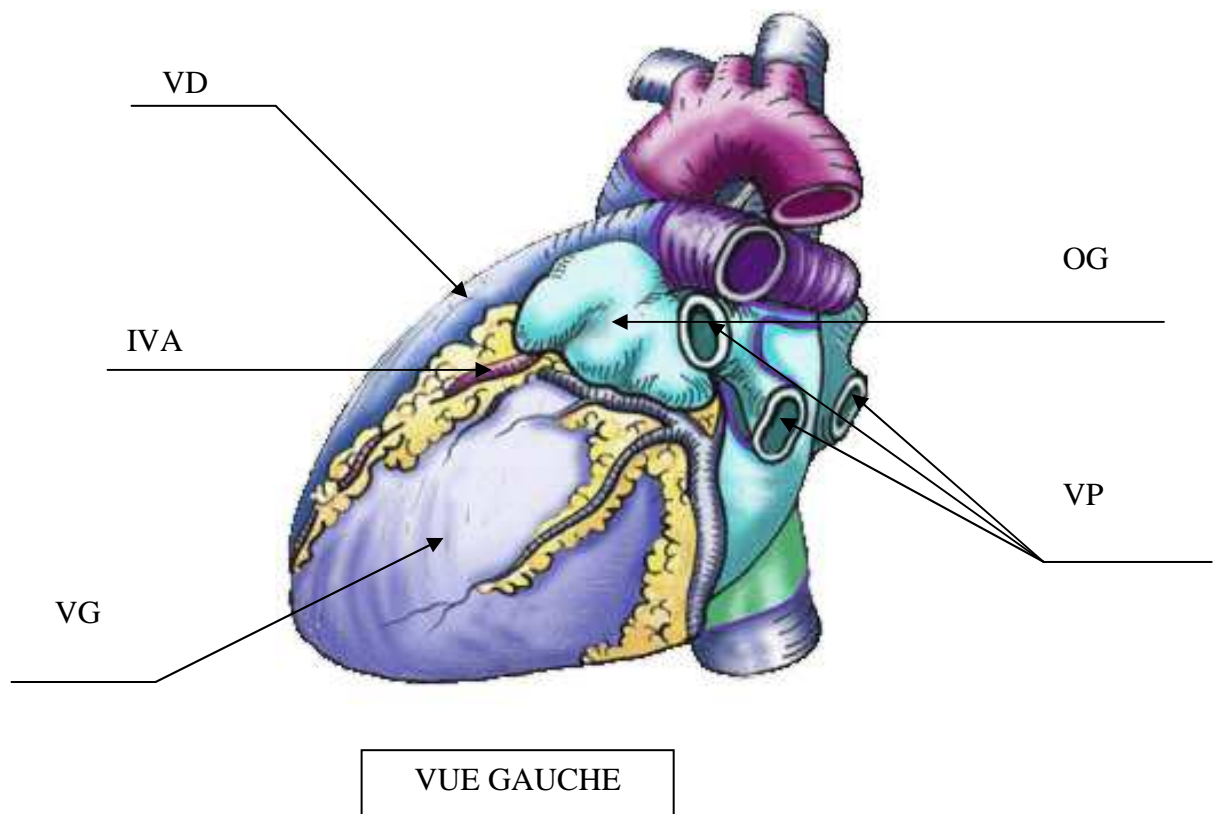
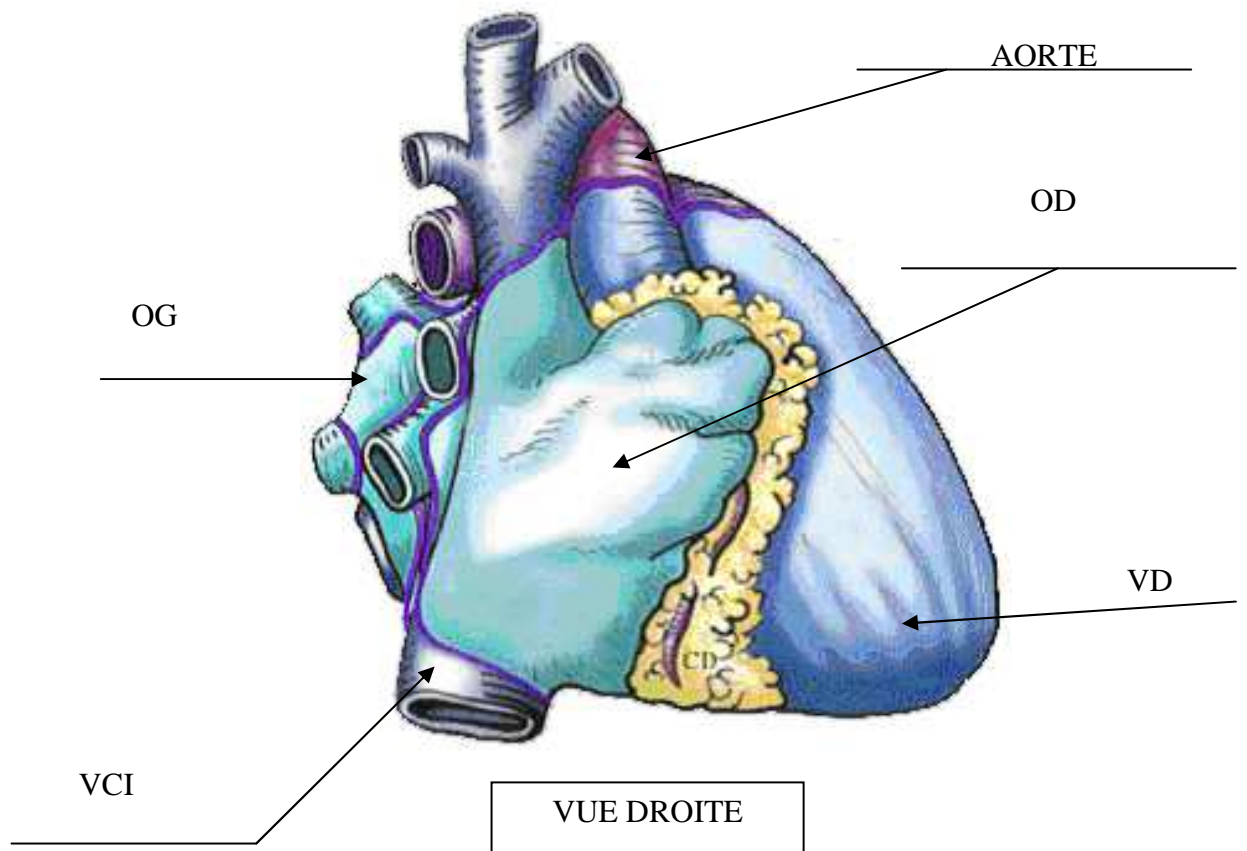
10-11 Vous êtes infirmier(e) aux urgences, vous recevez une femme enceinte de 32 semaines pour une entorse de cheville sans autre traumatisme. Cette patiente est allongée sur un brancard, son évaluation de la douleur (EVA) s'élève à 3. Au bout de 15 mn vous constatez une nette altération de la conscience de cette patiente.

...../ 10 pts






















a- Citez la cause la plus probable de ce malaise
Syndrome cave

b- Quelle mesure allez vous mettre en œuvre pour y remédier ?
Mise en place d'un coussin sous la fesse droite

10-14 Légendez les schémas ci-dessous.
...../ 10 pts



- a. Parmi les sept possibilités de contrôles pré-transfusionnels proposés ci-dessous, indiquez les groupes A, B, AB et O dans les cases de la colonne « GROUPES ».
...../ 10 pts

GROUPES	Sérum anti-A	Sérum anti-B	Sérum anti-AB
AB			
			
A			
B			
			
O			
			

- b. Vous disposez d'un mélange de 200 ml d'une substance pure X et de 0,8 litre d'eau.
Sans faire apparaître vos calculs :
...../ 10 pts

- a- Quelle sera la concentration de la substance X (en %) ?
20%

Vous prélevez 10% de ce mélange auquel vous ajoutez 0,4 litre d'eau

- b- Quelle sera la concentration de la substance X obtenue (en %) ?
4%

10-19 Vous devez administrer à un patient de la Dobutamine à la dose de $15\mu\text{g/kg/min}$.
...../ 10 pts

Le patient pèse 70 kg

La Dobutamine se présente en flacon de 20 ml contenant 250 mg

- Vous disposez de 2 ampoules de sérum physiologique de 20 ml, et d'une seringue de 50 ml.

Sans faire apparaître vos calculs, calculez le débit à sélectionner sur le pousse seringue (exprimé en ml/h).

dobutrex (250mg) dans 50ml (ajout de 30ml de sérum physiologique aux 20ml de la solution contenue dans le flacon)

Dose à perfuser en une heure :

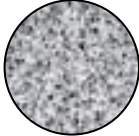

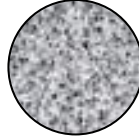
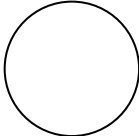
$$15\mu\text{g} \times 70 \times 60 = 63000\mu\text{g/h} \text{ soit } 63\text{mg/h}$$

250mg dans 50ml

$$63\text{mg} \text{ dans } X \text{ ml soit : } (63 \times 50)/250 = 12,6\text{ml}$$

le débit à afficher sur le pousse seringue sera de 12,6ml/h

10-20 Lors du contrôle pré-transfusionnel ultime au lit d'un patient, vous obtenez les résultats suivants :
.../ 10 pts

	Sang du patient	Sang du culot
Sérum ANTI-A		
Sérum ANTI-B		

e) Quel est le groupe sanguin du patient ?

AB

f) Quel est le groupe sanguin du culot ?

A

g) La transfusion est-elle possible ?

Oui

h) Pourquoi ?

Le culot (groupe A) ne possède pas d'antigène risquant d'agglutiner avec le sang du patient (groupe AB)

TABLE DES MATIERES

Remerciements.....	2
Sommaire	4
Introduction.....	6
Partie 1	7
I. Un contexte et une idée de départ inscrite dans le champ conceptuel de la docimologie.	7
1. Le terrain de l'étude : un établissement de formation professionnelle.	7
1.1. L'observateur.....	7
1.2. Un établissement de formation et une profession en évolution.....	9
1.3. Une évaluation en devenir	11
1.4. L'Infirmier Anesthésiste et sa formation dans le Contexte Européen.....	12
1.4.1. La formation des IADE dans le cadre des accords de Bologne	13
1.4.2. Les IADE en Europe	14
1.4.3. La diversité des rôles et des formations des IADE en Europe.....	14
1.5. Un projet européen pour les étudiants infirmiers anesthésistes ?.....	15
1.6. L'examen à l'entrée à l'origine de la question de départ	18
1.6.1. Les épreuves d'admission : examens, concours d'entrée ou examens classants ?	19
1.6.2. Les modalités de l'examen classant à l'entrée	20
1.7. L'origine de la question de départ.....	21
2. La docimologie, une science de l'éducation et de la formation issue de la psychologie expérimentale : présentation du concept et notions associés.....	23
2.1. Science de la notation, oui, mais pas seulement.....	23
2.2. Phase une : Naissance de la docimologie, le temps des constats	23
2.2.1. Les constats : L'Enquête CARNEGIE, un pavé dans la mare.	25
2.2.2. A ce propos, l'examineur se connaît-il lui-même ?	29
2.2.3. L'illusion de la double correction.	30
2.2.4. Conclusion.....	31
2.3. Phase deux : Du constat à ses causes, l'évaluation de la note et ses biais.....	31
2.3.1. Notion d'effet de tendance centrale	32
2.3.2. Notion d'effet de stéréotypie.....	33
2.3.3. Notion d'effet ROSENTHAL ou « effet PYGMALION ».....	33
2.3.4. Notion d'effet de contraste (ou effet de comparaison ou effet d'ordre de correction).....	35
2.3.5. Notion d'effet de halo	37
2.3.6. L'arrivée des sociologues et la mise en évidence de la notion de biais sociaux	37
2.4. Phase trois : un retour aux fondamentaux et l'amorce de propositions d'améliorations	39
2.4.1. Le commentaire à la note : une notion à améliorer.....	39
2.4.2. Bruno SUCHAUT :Une synthèse du retour aux fondamentaux	39

2.4.3.	La « constante macabre »	41
2.4.4.	En résumé de cette troisième phase	42
2.5.	En conclusion : Qu'en est-il aujourd'hui, de la docimologie ?	42
Partie 2	45
II. Problématisation de la question de départ : étude détaillée de l'examen et compléments conceptuels.	45
1.	Le constat de départ : une question d'efficacité de la sélection.....	45
1.1.	Un retour aux constats docimologiques historiques	45
2.	« Les premiers seront les derniers » ?	46
2.1.	La notion de rang relatif à l'admission.....	47
2.2.	L'échantillon étudié, le type d'étude et les outils utilisés	48
2.3.	Présentation descriptive graphique préliminaire	49
2.4.	Analyse statistique et correction méthodologique.....	51
2.5.	En conclusion : sommes nous fiables en terme d'évaluation initiale ?	54
3.	L'examen et les questions qui le compose : la notion de sélection et le concept associé de discrimination	54
3.1.	De la discrimination	55
3.1.1.	Les définitions générales.....	55
3.1.2.	Des confusions qui renforcent l'aspect péjoratif.....	56
3.1.3.	Sélection ou discrimination ?	58
3.2.	La courbe de GAUSS : un culte, pas une fatalité	58
4.	Un candidat, un jury et des épreuves :Les trois composantes de l'examen d'entrée... ..	61
4.1.	Le candidat	61
4.2.	Le jury	63
4.2.1.	Les forces en présence	64
4.2.2.	La responsabilité du jury	66
4.2.3.	Le rôle du jury	68
4.3.	Les épreuves	70
4.3.1.	La distribution des notes finales.....	71
4.3.2.	L'épreuve écrite : une distribution trop gaussienne et peu de poids de dans la décision finale	72
4.3.3.	La distribution et le poids déterminant des notes de l'épreuve orale dans la décision finale	75
4.3.4.	Il n'existe aucun lien entre l'épreuve écrite et l'épreuve orale	77
5.	Formulation de la problématique	78
Partie 3	81
III. La recherche : une étude docimologique portant sur le pouvoir discriminant des items de l'épreuve écrite	81
1.	Les buts de l'étude.	81
2.	L'échantillon exploité et les méthodes utilisées	81
2.1.	Un échantillon représentant sept années de résultats à l'examen d'entrée.....	82
2.2.	La méthodologie de l'étude	82
2.2.1.	Le choix de l' indice de discrimination.....	83
2.2.2.	Le choix de l'indice de difficulté	86
3.	Résultats de l'étude	87
3.1.	Pour faciliter la lecture : un exemple de la méthode concernant le repérage des indices de discrimination et de difficulté	87

3.1.1. Une question difficile n'est pas une question discriminante apte à la sélection.....	88
3.1.2. Plus le nombre des questions discriminantes est important, plus l'épreuve est sélective.....	89
3.2. Extraction des vingt deux questions discriminantes mises en évidence.....	92
3.3. Une référence nécessaire de classification des questions.....	96
4. Discussion :les QCM et les CROC sont les plus performants.....	97
4.1. Un exemple de question discriminante par types retenus.....	98
4.2. Analyse de la performance implicite des typologies des questions retenues ..	102
5. Critique et limites et de l'étude.....	105
6. conclusion	105
Partie 4.....	107
IV. Propositions d'améliorations et possibilités de réinvestissement professionnel...	107
1. Propositions d'amélioration de l'épreuve écrite	107
1.1. Propositions stratégiques	107
1.2. Propositions d'ingénierie de conception	110
2. Réinvestissement professionnel de l'étude	112
3. Conclusion générale.....	113
Références et Index.....	115
Table des Figures	115
Références bibliographiques.....	116
ANNEXES	123
I. Annexe 1.....	124
II. Annexe 2.....	125
III. Annexe 3.....	126
IV. Annexe 4.....	129
V. Annexe 5.....	130
VI. Annexe 6.....	146
Table des Matières	164



The discriminating power of the written tests during an entrance examination,

Research of improvements through a docimological approach.

In France, the importance of the entrance exam for training institute is major. The value of the grade given through this exam remains what is most important. This grade is double-edged and aims at giving the best appraisal of the candidate and thus avoiding some kind of gray zone among candidates that is to say uncertainty from the jury.

Since nearly eighty years, docimology endeavors to track down biased corrections. Lately, its extension concerning the measurement of the quality of learning attempts to qualify and quantify subjects (also called items) given during the tests.

Through the study of question given during the written exam, this study is intended to clarify the technical elements that may improve the quality of this type of examination. Finding questions that are best able to select clearly the eligible candidates and those who are not eligible, is a major point in this study.

The aim of this work is twofold: to clarify the eligibility status by avoiding the notes near average and, at the same time, to lessen the feeling of frustration of ineligible candidates who obtained a result very close to the limit-grade considered for the entrance exam.

The method used consists in giving a contextual and conceptual approach of the entrance examination in a school of Nurse-Anesthetists, completed by a technical study of the methods of the written test.

The conclusion of this study gives suggestions for improvements of the strategies and the mechanism of the very process of designing the written entrance exams.

Keywords: ranking Review, Evaluation, written Proof of eligibility, Discrimination index, selection.

Author : Philippe PELLETIER

Supervisor : Josette LAYEC

Le pouvoir discriminant des épreuves écrites lors d'un examen d'entrée classant :

Recherche d'amélioration par une approche docimologique.

En France, l'importance des concours d'entrée en institut de formation est majeure. A travers eux, la valeur de la note attribuée reste l'élément de référence. Tantôt sésame, tantôt couperet, cette note se doit de représenter clairement ces deux orientations afin d'éviter une zone de flou génératrice d'incertitude tant chez les candidats que chez les jurys.

Depuis près de 80 ans, la docimologie s'attache à débusquer les biais de correction. Plus récemment son prolongement éduométrique tente de qualifier et de quantifier les questions (également nommées items) posées lors des épreuves.

A travers l'étude des questions écrites proposées lors d'un examen classant, ce travail tente de clarifier les éléments techniques susceptibles d'améliorer la qualité de ce type d'épreuve. Le cas des questions discriminantes, les plus à même de sélectionner de façon claire les candidats admissibles de ceux qui ne le sont pas, est particulièrement abordé.

L'objectif de ce travail est double : clarifier le statut d'admissibilité en évitant les notes proche de la moyenne, et limiter le sentiment de frustration des candidats non admissibles ayant obtenus un résultat trop proche de la limite considérée pour l'examen.

La méthode retenue consiste en une approche contextuelle et conceptuelle d'un examen d'entrée dans une école d'Infirmiers(ères) Anesthésistes, complétée par une étude docimologique de type éduométrique concernant l'épreuve écrite.

La conclusion de cette recherche présente les propositions d'améliorations à apporter dans les champs de la stratégie et de l'ingénierie de conception d'une épreuve écrite d'admissibilité.

Mots clés : Examen classant, Docimologie, Epreuve d'admissibilité écrite, Indice de discrimination, Sélection.

Auteur : Philippe PELLETIER

Directeur de mémoire : Josette LAYEC