

Ecole Nationale Supérieure de Formation de l'Enseignement Agricole



**Master 2**

« Métiers de l'Enseignement, de l'Education et de la Formation »

Enseignant du Second Degré

## Mémoire

**Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».**

*Etude et comparaison des notes attribuées en fonction de la position de la copie dans 56 séries de copies issues des mathématiques, physique-chimie, français et philosophie.*

Amélie COMBELLES

Jury :

Yvan ABERNOT, ENSFEA : Directeur de mémoire

Dominique BROUSSAL, Maître de conférence, Université Jean-Jaurès Toulouse : Examineur

Sylvie FERNANDES, Docteur en sciences de l'éducation, ENSFEA : Examineur

Mai 2017



## SOMMAIRE :

<b>INTRODUCTION .....</b>	<b>2</b>
<b>1. CADRE THEORIQUE ET QUESTIONNEMENT .....</b>	<b>3</b>
<b>2. RECUEIL ET TRAITEMENT DES DONNEES DE TERRAIN .....</b>	<b>20</b>
<b>3. ANALYSE ET DISCUSSION DES RESULTATS DE L'ETUDE .....</b>	<b>25</b>
<b>PERSPECTIVES DES SUITES DE RECHERCHE .....</b>	<b>32</b>
<b>LIMITES ET DIFFICULTES RENCONTREES DURANT CE TRAVAIL .....</b>	<b>33</b>
<b>CONCLUSION : PERSPECTIVES D'EXPLOITATION PROFESSIONNELLE .....</b>	<b>34</b>
<b>BIBLIOGRAPHIE .....</b>	<b>36</b>
<b>SITOGRAPHIE .....</b>	<b>37</b>
<b>TABLE DES FIGURES .....</b>	<b>38</b>
<b>TABLE DES TABLEAUX .....</b>	<b>38</b>
<b>TABLE DES MATIERES .....</b>	<b>40</b>
<b>ANNEXES</b>	

## INTRODUCTION

*« Si l'écolier, l'élève, puis l'étudiant, parvient à éloigner de sa mémoire les arides tourments que lui ont fait connaître l'apprentissage de la vie à travers l'instruction (clef pour un avenir digne de ce nom), il lui reste souvent le vif ressentiment de l'évaluation, de la sanction, en un mot : de LA NOTE. »* Mireille Coron (mémoire de DEA en Sciences de l'Education, 1990).

*« L'évaluation est trop subjective, les points sont attribués n'importe comment... », « Si je passe après lui, je vais encore avoir une mauvaise note... ».* Voilà un extrait des dires des élèves.

Partant de ces constats, à l'aube de mon entrée dans le professorat, j'ai décidé de lever le voile sur ce sujet, qui peut être parfois, « tabou » : l'évaluation est-elle réellement objective ?

Ne pouvant pas traiter l'entière étendue de cette question (pourtant forte intéressante), le travail de recherche présenté dans ce mémoire porte sur la validité de l'évaluation et plus particulièrement sur les effets d'ordre ou de position rencontrés lors de la correction.

Depuis le milieu du 20ème siècle, plusieurs scientifiques ont démontré la présence et l'existence de ces « effets parasites », mais dans ce travail recherche, l'idée innovante développée repose sur la comparaison de l'influence de l'effet d'ordre entre les disciplines dites « normées » (types mathématiques ou physique-chimie) et les disciplines dites « moins normées » (types français ou philosophie).

Pour ce faire, la présente étude s'articulera en trois parties. La première partie sera consacrée à la présentation du cadre théorique bordant notre question de recherche et à l'étude des éléments bibliographiques associés.

La seconde partie s'attachera à décrire scrupuleusement les protocoles adoptés pour recueillir, traiter et analyser les données expérimentales.

Enfin, la troisième et dernière partie sera composée de l'analyse et de la discussion des résultats obtenus.

## 1. CADRE THEORIQUE ET QUESTIONNEMENT

---

### 1.1 CONTEXTE

#### 1.1.1 L'EVALUATION, PILIER DU SYSTEME EDUCATIF FRANÇAIS :

- **Qu'est-ce que l'évaluation ?**

Du côté des spécialistes des Sciences de l'éducation il y a une pluralité de définitions, toutes très intéressantes, mais pour cette étude nous n'en retiendrons qu'une :

*« L'évaluation consiste en une mesure ou une appréciation, à l'aide de critères, de l'atteinte d'objectifs ou du degré de proximité d'une production par rapport à une norme » Abernot, 1996.*

Dans le cadre scolaire, l'évaluation c'est aussi donner *une valeur* aux productions des élèves. Cette valeur peut être donnée sous différentes formes : appréciation, échelle numérique ou ordinale...

La transposition « production de l'élève » (écrit, oral, épreuve pratique...) en une valeur est un mécanisme très important. Cela permet de mettre l'ensemble des productions scolaires (qui ne sont pas forcément comparable entre elles dans leurs domaines d'origine) dans un même système de valeur et de pouvoir ainsi réaliser des opérations (additions, moyennes, comparaisons, classements...) à différentes échelles (intra discipline, inter discipline, par élève, par classe, par établissement...).

Le plus souvent, comme le souligne à juste titre Y. Abernot (1996), la valeur d'une production scolaire prend la forme d'une note. Ce constat peut soulever plusieurs réactions notamment des questions concernant la note, sa pertinence et son utilité. Nous allons en rester là pour l'instant car nous reviendrons sur ces interrogations ultérieurement.

Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

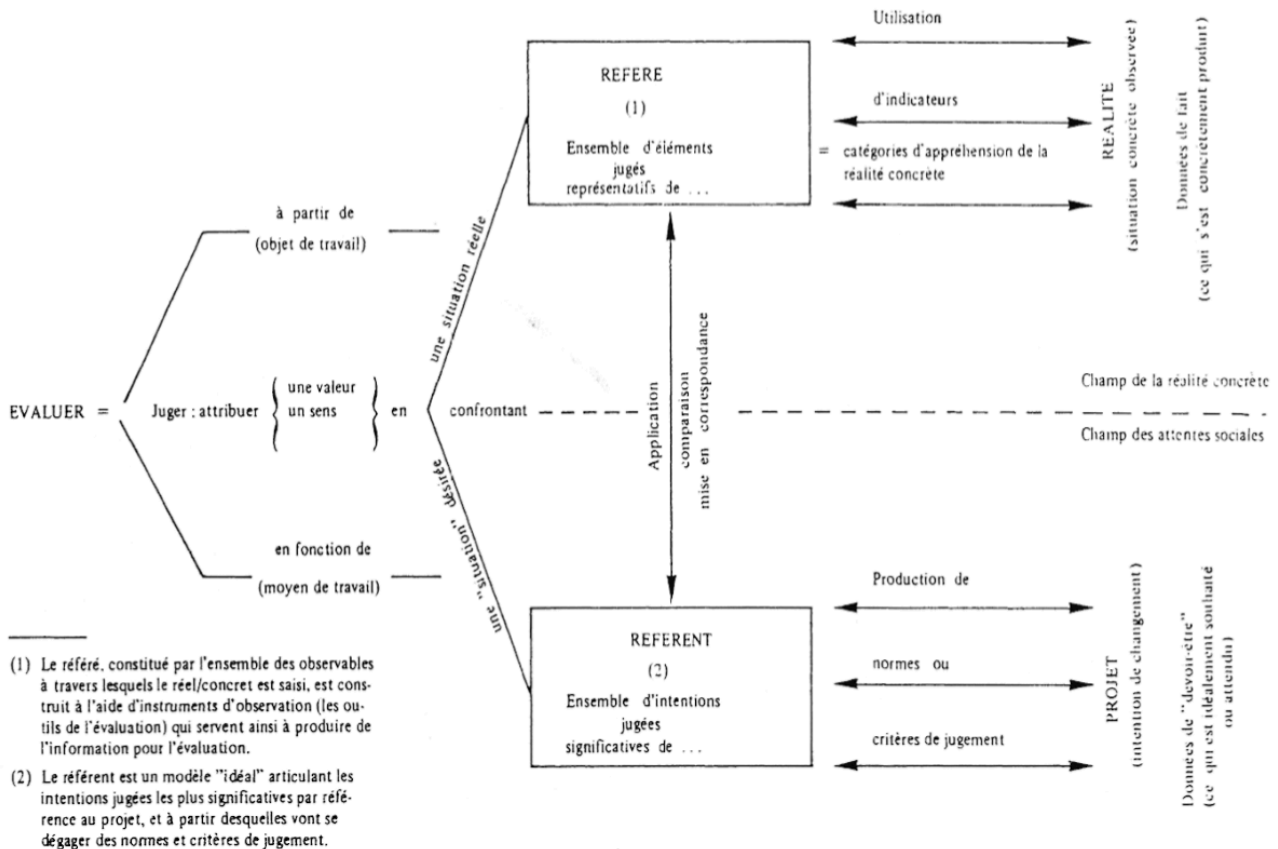


Figure 1  
La double articulation dans l'opération d'évaluation.

Le schéma ci-dessus proposé par Charles Hadji dans « L'évaluation, règles du jeu des intentions aux outils. » (1989) modélise parfaitement la définition précédente d'Y. Abernot.

En effet, lors de l'évaluation, l'enseignant(e) évalue une performance en confrontant une situation réelle (une production d'élève, une attitude, un savoir-faire) à une situation désirée (la norme, un idéal, la production d'un autre élève...) sans perdre de vue la visée finale : évaluer la (ou les) compétence(s) de l'élève.

La performance jugée lors d'une évaluation repose sur une partie infime de l'enseignement total réalisé en classe mais c'est sur cette dernière que l'enseignant(e) se base pour déterminer si un élève a atteint ou non les objectifs fixés pour cette séquence. C'est pourquoi il est primordial de bien fixer les critères et les buts de l'évaluation.

- **Pourquoi, pour qui, quand et comment évaluer ?**

L'évaluation est une pratique centrale de l'acte d'enseignement, chaque enseignant(e) la pratique quotidiennement. Pour ces derniers, c'est avant tout une pratique « régulatrice ». Elle leur permet de juger leurs actions d'enseignement, voir où les élèves se situent par rapport aux objectifs visés et ainsi adapter ou améliorer leurs pratiques pédagogiques afin de toujours optimiser la progression des élèves.

Pour M. Musial, F. Pradère et A. Tricot, cette visée « régulatrice » n'est pas la seule lorsque l'on parle d'évaluation. En effet, d'après eux, l'évaluation peut avoir des visées très différentes qui vont de la plus légitime aux plus discutables. Dans leur ouvrage « *Comment concevoir un enseignement ?* » ils en citent neuf :

- Sélectionner : lors d'un concours par exemple. Prendre N candidats sur les X qui ont passé le concours.
- Valider : on évalue pour vérifier que le but visé a été atteint.
- Sanctionner : on évalue pour punir les élèves ou un élève. L'évaluation est ici utilisée comme menace, notamment l'évaluation « surprise ». Bien entendu cette utilisation est une perversion de l'évaluation et est prohibée dans la plupart des institutions scolaires.
- Motiver : Lorsque que l'on focalise l'évaluation sur les progrès réalisés par l'élève, elle peut être utilisée comme moteur d'apprentissage.
- Aider à apprendre : on évalue pour fournir à l'élève une aide précise et personnalisée à son apprentissage. En situant sa production par rapport à une référence, on peut dire précisément à l'élève les étapes qu'il doit encore franchir (identification claire des connaissances manquantes...). Avec la visée « régulatrice », c'est là une des principales utilités de l'évaluation.
- Prendre conscience : A l'inverse de la précédente, dans cette situation, on évalue pour montrer à l'élève qu'il est capable de progresser en mettant en avant seulement les acquis. Ici l'évaluation a surtout pour objectif de redonner confiance en soi à l'élève. D'ailleurs cette méthode est souvent utilisée avec des élèves en grandes difficultés, proche du décrochage scolaire.
- Dépister ou repérer : Evaluer en suivant une méthodologie standardisée afin d'identifier une population à risque concernant les troubles de l'apprentissage.

Les familles des élèves identifiés seront alors contactées et invitées à se rapprocher d'un spécialiste afin de faire des examens plus approfondis et de poser un diagnostic.

- Diagnostiquer : Evaluer pour faire l' « état des lieux » des acquis des élèves. Repérer les points forts sur lesquels l'enseignant va pouvoir s'appuyer (ancres) pour apporter de nouveaux apprentissages et repérer aussi les difficultés qu'ils rencontrent afin d'y remédier.
- Evaluer le système : Evaluation du type étude PISA (évaluation du système éducatif à l'échelle nationale et comparaison entre différents pays).

Nous pouvons constater que les divers objectifs visés par l'évaluation sont de nature bien différente. Si nous les observons de plus près, nous nous rendons compte que les résultats qui en découlent ne servent pas les mêmes acteurs. Par exemple, les résultats d'une évaluation internationale (type PISA) ne sont pas le principal centre d'intérêt des élèves alors que le résultat des leurs évaluations sommatives l'est beaucoup plus.

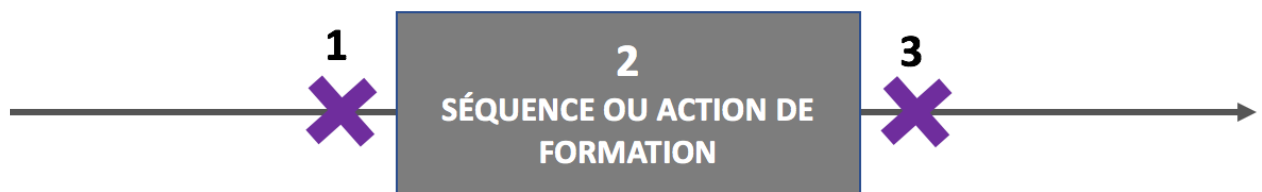
Ainsi, nous pouvons identifier quatre types d'acteurs concernés par l'évaluation : les enseignants (voire l'ensemble de l'équipe éducative), les élèves, les familles et le système et/ou la société.

En fonction de l'objectif visé (par exemple : sélectionner ou diagnostiquer), la méthode utilisée pour évaluer et le moment pour réaliser l'évaluation ne seront pas les mêmes (et inversement). Le tableau (voir page suivante) présente les différents types d'évaluations, leurs visées respectives et les démarches utilisées par rapport à la séquence de formation. Pour réaliser ce tableau, je me suis appuyée sur celui proposé par C. Hadji dans son ouvrage « L'évaluation, règles du jeu : des intentions aux outils » en y intégrant les fonctions et les démarches des différents types d'évaluations énoncées par JM. De Ketele.

Ainsi, nous pouvons constater qu'avant la séquence d'enseignement (1) C. Hadji (1989) et JM. De Ketele (1993) nous recommandent de réaliser seulement des évaluations diagnostiques, pronostiques ou prédictives qui ont pour but de préparer l'action de formation à venir en s'intéressant aux élèves et à leurs différentes caractéristiques (prés-acquis ou ancre, fausse idée à déconstruire, participation active et/ou besoin d'être sollicité...).

Dans un second temps, tout au long de la séquence de formation (2), ils nous invitent à réaliser des évaluations formatives ou progressives afin d'aider les élèves à progresser en identifiant avec eux, d'après leurs activités (exercices/raisonnement), leurs acquis et/ou les processus de réflexion positifs mis en œuvre.

Pour finir, à l'issue de la séquence de formation (3), les deux spécialistes nous conseillent de réaliser des évaluations sommatives afin de valider ou de certifier les capacités acquises par les élèves durant la séquence de formation passée en se basant sur leurs productions.



AVANT L'ACTION DE FORMATION (1)	PENDANT L'ACTION DE FORMATION (2)	APRÈS L'ACTION DE FORMATION (3)
<p><b>TYPE D'ÉVALUATION:</b></p> <ul style="list-style-type: none"> <li>• DIAGNOSTIQUE</li> <li>• PRONOSTIQUE</li> <li>• PRÉDICTIVE</li> </ul> <p><b>FONCTION:</b></p> <ul style="list-style-type: none"> <li>• D'ORIENTATION (PRÉPARER UNE NOUVELLE ACTION)</li> </ul> <p><b>DÉMARCHE:</b></p> <ul style="list-style-type: none"> <li>• INTERPRÉTATIVE (DONNER DU SENS À UN ENSEMBLE D'INDICES QUANTITATIFS OU QUALITATIFS)</li> </ul> <p><b>CENTRÉE SUR:</b></p> <ul style="list-style-type: none"> <li>• LES PRODUCTEURS (ÉLÈVES) ET LEURS CARACTÉRISTIQUES</li> </ul>	<p><b>TYPE D'ÉVALUATION:</b></p> <ul style="list-style-type: none"> <li>• FORMATIVE</li> <li>• « PROGRESSIVE »</li> </ul> <p><b>FONCTION:</b></p> <ul style="list-style-type: none"> <li>• FORMATIVE</li> </ul> <p><b>DÉMARCHE:</b></p> <ul style="list-style-type: none"> <li>• DESCRIPTIVE (IDENTIFICATION ET DESCRIPTION DES ACQUIS ET DES PROCESSUS POSITIFS)</li> </ul> <p><b>CENTRÉE SUR:</b></p> <ul style="list-style-type: none"> <li>• LES PROCESSUS</li> <li>• L'ACTIVITÉ</li> </ul> <p style="text-align: right;">} DE PRODUCTION</p>	<p><b>TYPE D'ÉVALUATION:</b></p> <ul style="list-style-type: none"> <li>• SOMMATIVE</li> <li>• TERMINALE</li> </ul> <p><b>FONCTION:</b></p> <ul style="list-style-type: none"> <li>• CERTIFICATIVE (CERTIFIER/VALIDER)</li> </ul> <p><b>DÉMARCHE:</b></p> <ul style="list-style-type: none"> <li>• SOMMATIVE</li> <li>• (FAIRE LA SOMME DES ACQUIS)</li> </ul> <p><b>CENTRÉE SUR:</b></p> <ul style="list-style-type: none"> <li>• LES PRODUITS</li> </ul>

**Tableau 1 : Fonctions, types et démarches de l'évaluation selon sa place par rapport à la séquence de formation.**



- **La place de l'évaluation dans le cadre réglementaire : La note une obligation légale ou une attente sociale ?**

Si l'on se réfère au code de l'éducation, nous pouvons constater que l'acte évaluatif est une obligation de service à laquelle tout enseignant(e) doit se conformer, quel que soit le niveau dans lequel il exerce (de l'école primaire à l'enseignement supérieur).

*La scolarité est organisée en cycles pour lesquels sont définis des objectifs et des programmes nationaux de formation comportant une progression régulière ainsi que des critères d'évaluation. (...) Dans l'enseignement primaire, l'évaluation sert à mesurer la progression de l'acquisition des compétences et des connaissances de chaque élève. Cette logique d'évaluation est aussi encouragée dans l'enseignement secondaire. Article L311-1 du code de l'éducation.*

Le système éducatif français repose sur ces évaluations omniprésentes. Nous avons vu précédemment qu'elles sont indispensables aux enseignants car elles sont l'instrument de mesure de la progression des élèves mais elles garantissent aussi un « résultat », une preuve du travail effectué par l'école aux tiers (familles et société tout entière).

*Durant la scolarité, l'appréciation de l'acquisition progressive des connaissances et des compétences s'exerce par un contrôle continu assuré par les enseignants sous la responsabilité du directeur ou du chef d'établissement. Article L311-7 du code de l'éducation.*

*Les parents sont tenus régulièrement informés de l'évolution des acquis scolaires de leurs enfants (...) Article D111-3 du Code de l'éducation*

*Les enseignants tiennent informés les parents d'élèves et les aident à suivre la scolarité de leurs enfants. Article L912-1 du Code de l'éducation.*

En revanche, si chaque enseignant(e) se doit d'évaluer ses élèves, aucun texte réglementaire ne le contraint à une forme particulière d'évaluation ni de transposition des résultats (hors CCF). Tout cela relève de son propre arbitrage : de sa liberté pédagogique.

*Les enseignants sont responsables de l'ensemble des activités scolaires des élèves. (...)*

*Les enseignants apportent une aide au travail personnel des élèves et en assurent le suivi. Ils procèdent à leur évaluation. (...) Article L912-1 du Code de l'éducation.*

*La liberté pédagogique de l'enseignant s'exerce dans le respect des programmes et des instructions du ministre chargé de l'éducation nationale et dans le cadre du projet d'école ou d'établissement avec le conseil et sous le contrôle des membres des corps d'inspection.*

*Le conseil pédagogique (...) ne peut porter atteinte à cette liberté. Article L912-1-1 du code de l'éducation.*

Par convention, nous avons l'habitude de convertir le résultat d'une évaluation en une note numérique (échelle de 21 degrés – de 0 à 20 -) mais cette forme de notation n'a en rien un caractère obligatoire. D'ailleurs, d'autres systèmes de notation communément utilisés dans des pays étrangers (type échelle ordinaire A, B, C, D, E, F pour l'Angleterre ou échelle de 5 degrés – 5 à 1 – en Allemagne) sont utilisés de façon régulière par des enseignants français qui les jugent plus pertinents que l'échelle de notation de 21 degrés.

Il serait intéressant de se pencher sur les raisons qui font que certains enseignants se désintéressent du système de notation sur 21 degrés et choisissent d'adopter un autre système de notation, quitte à se marginaliser et donc à complexifier leur travail au sein d'une équipe pédagogique par la justification permanente de leur choix ou par la recherche continue d'une échelle de conversion (entre le système de notation utilisé par l'enseignant et l'échelle de 21 degrés « conventionnel ») afin d'être compris par tous, par exemple.

Ce constat relance le débat autour de la « valeur » de la note, de sa fiabilité, de ses représentations et de ses incidences sur les élèves.

Effectivement, la note peut avoir des effets négatifs tant sur la gestion générale de la classe que sur la motivation individuelle de chaque élève. En effet, elle creuse le fossé de l'hétérogénéité à cause de la compétition et des classements qu'elle engendre. Ainsi, les élèves en réussite entrent dans une compétition positive et motivante pour maintenir leurs résultats et surtout leur position dans le classement ; Alors que les élèves en difficultés entrent dans une spirale négative et démotivante.

De plus, l'utilisation de la note lors des évaluations formatives biaise l'objectif recherché. En effet, elle fait « écran » entre le savoir, les acquis de l'élève et sa marge de progression pour atteindre l'ensemble des objectifs fixés pour la séquence. Dans ce cas là aussi, ce sont les élèves les plus en difficultés qui seront stigmatisés par les « mauvaises notes » et donc incités à relâcher leurs efforts. Dans un article publié en 2012 dans *Regard croisés sur l'économie*, Pierre Merle fait un lien entre « estime de soi et notation », il constate qu'utiliser la note lors de chaque évaluation (quel qu'en soit le type) engendre un fort stress chez les élèves qu'il nomme « pression évaluative » et va même jusqu'à affirmer que le système de notation utilisé en France est une des causes du décrochage scolaire.

En revanche, la note prend tout son sens lors des évaluations sommatives de fin de séquence, qui visent (comme indiqué dans le tableau 1 page 6) à valider les savoirs acquis durant une ou plusieurs séquences.

En outre, la note est un indicateur adéquat pour informer les familles des résultats des élèves, à condition qu'elle soit accompagnée d'un commentaire ou d'une appréciation permettant une bonne interprétation de cette dernière.

La plupart des spécialistes des sciences de l'éducation s'accordent pour dire que pour être efficace, la note doit seulement être utilisée lors des évaluations sommatives qui ont pour but de valider ou de certifier les acquis des élèves.

Nous venons de déterminer les moments les plus propices à l'utilisation de la note lors de l'évaluation. A présent, nous allons nous intéresser à la validité, à la fidélité et à la sensibilité de ces notes afin de voir si elles sont réellement « justes ».

Pour nous aider dans cette tâche, nous allons nous appuyer sur les travaux docimologiques et notamment sur ceux réalisés par Henri Piéron, père fondateur de cette discipline.

### 1.1.2 LA DOCIMOLOGIE ET SES APPORTS SUR LA QUESTION DE L'ÉVALUATION

D'après Gilbert De Landsherre, « *La docimologie est une science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés* ».

C'est Henri Piéron qui, en 1929, créa le terme « docimologie » à partir du vocable grec : *dokimé* qui signifie « épreuves » et *logos* qui signifie « science ». Ce néologisme sera popularisé en 1963 lors de la parution de son ouvrage « Examens et docimologie » dans lequel, H Piéron pose les fondements de cette nouvelle discipline en présentant les multiples travaux réalisés sur cette problématique pendant la première moitié du 20<sup>ème</sup> siècle.

Dans ses débuts, la docimologie dite « classique » s'intéresse particulièrement à la fidélité, à la sensibilité et à la validité des examens. Certains spécialistes (comme G. De Landsherre (1980)) qualifient même la docimologie de cette période de science à caractère négatif qui met à jour les différents problèmes de fidélité, de validité et de sensibilité de l'évaluation sans proposer de solutions.

Par la suite, la docimologie entrera dans une phase « constructive » en tentant de proposer des méthodes et des techniques de notations afin de rendre plus juste les notes scolaires.

D'ailleurs, G. Noizet et JJ. Bonniol (1968-1969) créent un nouveau terme pour marquer cette progression dans la discipline de base. Ils écrivent : « *Si l'investigation docimologique dépasse le plan du constat, si elle permet un perfectionnement du système d'estimation par la connaissance expérimentale des mécanismes en jeu et des causes de distorsion dans le fonctionnement, alors la docimologie devient en fait une docinomie* ». Malheureusement, seul le terme docimologie a perduré, les autres nuances (dont la docinomie) ont périclité sauf peut-être pour quelques spécialistes...

Nous avons souligné précédemment les trois champs de travail de la docimologie classique : la fidélité, la sensibilité et la validité de l'évaluation. Nous allons à présent étudier plus en détails ces champs, en s'appuyant sur les différentes études réalisées portant sur les mathématiques, la physique-chimie, le français et la philosophie (les deux types de disciplines sur lesquelles porte mon expérimentation).

- **La fidélité**

Les docimologues disent d'une évaluation qu'elle est fidèle si les copies de la série corrigée obtiennent les mêmes notes quel que soit le correcteur ou la date de correction.

Depuis les débuts de la docimologie, plusieurs scientifiques se sont intéressés au critère de fidélité en évaluation. Cependant, nous allons nous focaliser sur les premiers travaux d'envergure réalisés sur cette question par deux scientifiques français : H. Laugier et D. Weinberg.

C'est dans le cadre de la célèbre enquête Carnegie (réalisée entre 1931 et la 2<sup>nd</sup>e guerre mondiale et impulsée par l'International Institute of Education de Colombia) visant à étudier spécifiquement les examens officiels de plusieurs pays que les deux scientifiques français se sont intéressés au baccalauréat.

Pour mener leurs recherches, ils prélevèrent au hasard (dans les archives de l'Office du Baccalauréat de l'Académie de Paris), 100 copies de :

- Composition de français,
- Version latine,
- Composition anglaise,
- Composition mathématiques,
- Composition physique,
- Dissertation de philosophie.

Ces dernières furent confiées, après avoir effacé toutes traces de corrections précédentes, à cinq correcteurs de chaque discipline habitués à corriger les épreuves du baccalauréat. Le tableau ci-dessous présente leurs résultats ; J'ai sciemment retiré les résultats de la version latine et de la composition anglaise afin d'observer uniquement les quatre disciplines choisies pour mon étude.

Epreuves/Disciplines	Ecart moyen	Ecart le plus fréquent	Ecart maximum
Composition française	3.29	6 et 7	13
Dissertation de philosophie	3.36	5 et 7	12
Composition mathématiques	2.05	4	9
Composition physique	1.83	4	5

**Tableau 2 : Ecart moyen, maximum et les plus fréquemment relevés entre les cinq correcteurs par H. Laugier et D. Weinberg lors de l'étude Carnegie.**

*Les scientifiques précisent qu'en aucun cas n'a été trouvée une note identique dans un couple de correcteurs.*

Nous pouvons constater que même si les écarts ne sont pas les mêmes en fonction des disciplines ou types de disciplines :

- Disciplines « normées » (mathématiques et physique) écart le plus fréquemment constaté : 4 points.
- Disciplines « moins normées » (français, philosophie) écart le plus fréquemment constaté : de 5 à 7 points.

L'ensemble des disciplines sont sujettes à des variations inter-individuelles (c'est-à-dire entre les différents correcteurs) lors de la notation.

Rappelons que l'étude porte sur des véritables copies du Baccalauréat, les notes obtenues ont donc une incidence réelle sur l'avenir des candidats (reçu ou ajourné).

Ainsi, à l'aide des données récoltées par H. Laugier et D. Weinberg, A. Agazzi a établi le bilan suivant :

	Ajournés par les 5 correcteurs	Admis par les 5 correcteurs	<b>Admis par les uns et ajournés par les autres</b>
Composition française	21 %	9 %	<b>70 %</b>
Dissertation philosophique	9 %	10 %	<b>81 %</b>
Composition mathématiques	44 %	20 %	<b>36 %</b>
Composition physique	37 %	13 %	<b>50%</b>

**Tableau 3 : Résultats des différents candidats (dont les copies ont été tirées au hasard) aux épreuves de français, philosophie, mathématiques et physique du baccalauréat en fonction des correcteurs.**

A la lecture de ce tableau, nous sommes en droit de nous demander si le facteur « chance » ne joue pas un grand rôle lors de l'obtention d'un diplôme type baccalauréat. En effet, nous constatons que le sort de 70 candidats sur 100 dépend du correcteur auquel il aura eu affaire pour l'épreuve de français. Ce ratio monte à 81% pour l'épreuve de philosophie et redescend pour les épreuves de sciences en passant à 50 % pour l'épreuve de physique et 36% pour l'épreuve de mathématiques.

Après avoir fait ces constats effarants, H. Laugier et D. Weinberg ont cherché à déterminer une « note vraie », note qui serait égale à la moyenne des notes données par un ensemble de correcteurs expérimentés dont l'effectif assurerait une stabilité de cette note.

Pour déterminer cet effectif, le couple Laugier Weinberg a utilisé la formule de Spearman-Brown en fixant le coefficient de fidélité à 0,99. Voici les résultats.

Composition mathématiques	13 correcteurs
Composition physique	16 correcteurs
Composition française	78 correcteurs
Dissertation philosophique	127 correcteurs

**Tableau 4 : Nombre de correcteurs nécessaires pour obtenir une « note vraie » ou stabiliser la note.**

Il faudrait donc 13 correcteurs pour stabiliser une note en mathématiques contre 127 en philosophie soit environ dix fois plus... Un tel constat laisse perplexe. Pour ne pas tirer de conclusion hâtive de ces faits, Y. Abernot (1996) nous rappelle que : « Il n'est pas du tout de semblable difficulté de traduire en note ce qui est déjà d'ordre quantitatif et de représenter par un nombre, ce qui est qualitatif ». Autrement dit, le respect du critère de fidélité n'est pas propre aux disciplines normées mais plutôt aux types d'objets évalués : objet quantitatif (type bonnes réponses dans un QCM) ou objet qualitatif (type argumentaire de dissertation). En effet, si la dictée était l'exercice réalisé lors de l'épreuve du baccalauréat de français, l'écart moyen entre l'ensemble des notes ou le nombre de correcteurs nécessaires pour obtenir la « note vraie » serait sûrement équivalent au nombre de correcteurs nécessaires pour stabiliser la note en mathématiques c'est-à-dire, une dizaine environ.

- **La sensibilité**

Pour qu'une évaluation soit « sensible », il faut que l'échelle de notation utilisée (ou le système de notation) soit adaptée à la complexité du produit à évaluer.

Nous avons vu précédemment que, par convention, la plupart des enseignants utilisent l'échelle de 21 degrés alors qu'elle n'est pas toujours pertinente. En effet, nous le constatons bien lorsque l'on observe des séries de notes qui oscillent toutes autour d'un faible intervalle par exemple de 8 à 12.

Pour les docimologues, il est possible de mesurer la sensibilité d'une évaluation en étudiant la dispersion des notes obtenues grâce à l'écart type.

Pour rappel, un écart type proche de 0 signifie que le correcteur attribue l'ensemble de ses notes dans un faible intervalle (de 8 à 12 par exemple). Dans ce cas-là, l'outil est faiblement sensible.

Un écart type proche de 5 signifie que le correcteur attribue autant de chacune des notes (deux 0, deux 1, ... , deux 20). Dans ce cas, l'outil est très sensible.

Un écart type proche de 10 signifie que le correcteur distribue seulement des notes extrêmes (beaucoup de notes entre 0 et 5 et entre 15 et 20 par exemple). Dans ce cas de figure, comme dans le premier, l'outil est faiblement sensible.

Le tableau ci-dessous présente l'écart type de la distribution des notes relevées dans le cadre de l'enquête Carnegie. Précisons que seuls les écarts-types des correcteurs « extrêmes » (celui ayant les notes les plus rapprochées et celui ayant la distribution la plus étendue) de chaque discipline sont présentés.

	Distribution concentrée	Distribution étendue
Composition française	2.91	3.85
Dissertation philosophique	2.22	4.25
Composition mathématiques	4	4.30
Composition physique	3.31	3.63

**Tableau 5 : Ecart-types observés sur des séries de notes relevées lors de l'enquête Carnegie.**  
*Tableau extrait d'Y. Abernot (1996)*

A la lecture de ce tableau, nous pouvons observer que les correcteurs des disciplines « normées » utilisent pratiquement tous de la même façon l'échelle de notation (faible écart entre le correcteur ayant la distribution la plus concentrée et celui ayant la distribution la plus étendue) et ils l'utilisent aussi dans sa globalité ; les écarts-types oscillant autour de 5, les notes distribuées doivent aller de 0 à 20 (sauf peut-être pour le correcteur de physique ayant la distribution la plus concentrée). En revanche, nous ne pouvons en dire autant pour les disciplines dites « moins normées ». On constate un écart considérable entre la distribution la plus étroite et la plus étendue ; 0.91 en français et 2,03 en philosophie. Cela traduit une utilisation différente de l'échelle de notation en fonction des correcteurs.



En outre, il est à noter que les distributions les plus étendues en français et en philosophie (ayant pour écarts-types : 3.85 et 4.25) s'apparentent aux écart types des disciplines « normées » (4.30 pour les mathématiques et 3.63 pour la physique). Cela signifie que certains correcteurs de français ou de philosophie utilisent l'ensemble de l'échelle de notation et distribue l'ensemble des notes dans les mêmes proportions (ou presque) tout comme des correcteurs de sciences.

- **La validité**

Lorsqu'il parle de la validité en évaluation, G. Delandsheere dit : « *Non seulement un test valide doit mesurer ce qu'il prétend mesurer, mais, en outre, il ne doit pas mesurer quelque chose d'autre.* » Ainsi, lors de la correction d'une évaluation aucune variable externe à la production ne devrait entrer en ligne de compte. Or, de nombreuses études démontrent qu'il est fort difficile d'attribuer une note qui reflète uniquement les caractéristiques intrinsèques d'une production. En effet, il existe des « effets parasites » qui influencent, le plus souvent inconsciemment, le correcteur au moment d'attribuer la note.

Grâce à de nombreuses recherches, les docimologues ont mis en évidence et ont nommé plusieurs types de biais : effets de contamination, effets de halo, effets de stéréotypie, effets d'ordre, de position et de contraste.

L'effet de contamination désigne l'influence que peut avoir l'avis des autres enseignants (via le dossier scolaire par exemple) sur le jugement du correcteur. J-P Caverni (1975) a démontré que cet effet est d'autant plus important que le niveau de la copie est élevé. C'est-à-dire qu'un dossier scolaire positif (comportant seulement des bonnes notes) ne fait pas augmenter significativement la note d'une copie médiocre. En revanche, un mauvais dossier scolaire (comportant seulement des notes médiocres) influence de façon significative la note d'un bon devoir.

L'effet de halo traduit l'influence de critères « affectifs » renvoyés par l'élève à son enseignant. Ce dernier sera par exemple sensible à une belle écriture sans fautes d'orthographe, à un comportement et une tenue en classe irréprochables, à une origine sociale particulière (...) et aura un préjugé favorable au moment de corriger la copie de cet élève « modèle » ce qui influencera positivement la note. La réciproque est tout à fait valable lorsque l'enseignant a un préjugé défavorable.

L'effet de stéréotypie est semblable à l'effet de contamination à la différence que, dans ce cas, ce n'est pas l'avis des confrères qui influence le jugement de l'enseignant mais plutôt le sien. En effet, l'enseignant sera influencé par les productions précédentes de l'élève. Pour illustrer ce « biais » G. Delandsheere (1980) relate l'anecdote suivante : « Un professeur de langue maternelle fait régulièrement des dictées. Bientôt, il connaît les élèves qui réussissent habituellement le mieux et le moins bien cet exercice. Si l'on détermine la fréquence des fautes « oubliées », non perçues par le correcteur, on constate que les oublis en faveur des bons élèves sont significativement plus élevés que pour les élèves faibles. Dans le premier cas, le maître s'attend à ne pas rencontrer d'erreurs ; dans le second il les guette. »

L'effet d'ordre ou de position, sujet central de ce travail de recherche, est un des biais de l'évaluation sur lequel JJ Bonniol a beaucoup travaillé. Il a d'ailleurs mené une étude expérimentale pour vérifier si oui ou non des effets d'ordre, constaté empiriquement par les apprenants, existent vraiment. Pour ce faire, il réunit deux groupes de neuf correcteurs afin de leur faire corriger une série de 26 copies de versions anglaises identiques à la différence que le premier groupe corrigea dans un sens et le second dans le sens inverse.

	Ordre direct (Copie 1 à copie 26)	Ordre indirect (Copie 26 à copie 1)
Première moitié	10.75	9.85
Deuxième moitié	9.24	10.14
Premier tiers	11.17	9.01
Second tiers	9.18	10.36

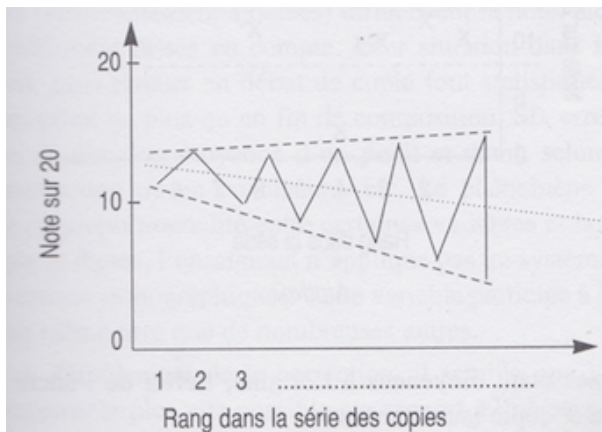
**Tableau 6 : Moyenne des notes attribuées par les 18 correcteurs, selon l'ordre de correction dans le cadre de l'étude de JJ. Bonniol (1965).**

*Tableau extrait de Noizet. G et Caverni. JP (1978) : pour chacune des 18 séries de notes, la moyenne a été ramenée à 10 (centrage) et l'écart type à 3 (réduction).*

En répétant cette expérience un nombre de fois suffisant pour juger les résultats représentatifs, JJ Bonniol fit deux constats majeurs :

- Les correcteurs sont plus sévères à la fin de la série qu'au début. Cela se traduit par la diminution de la moyenne des notes attribuées au fur et à mesure de la correction. Dans ce cas-là, nous parlons *d'effet d'ordre au sens strict*.

Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

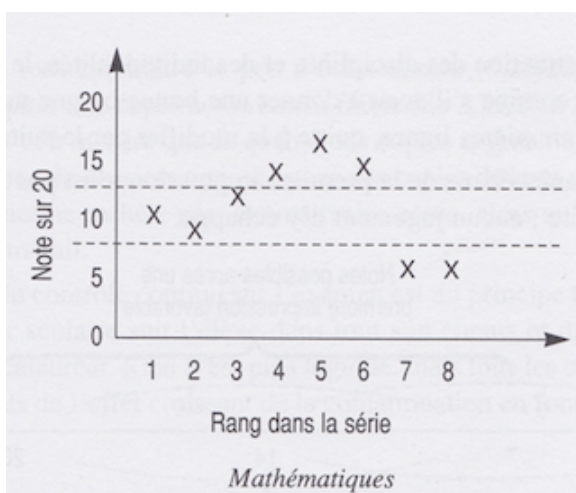


**Figure 2 : Modélisation des effets d'ordre au sens strict dans le processus de correction d'une série de copie.**

Y. Abernot (1996)

*La ligne de pointillés fins représente la diminution de la moyenne des notes attribuées au cours du processus de correction.*

- Les correcteurs notent par « contraste ». C'est-à-dire qu'ils sont influencés par les copies précédentes au moment d'attribuer une note, au moment de juger une copie. JJ Bonniol utilise le terme d'ancre pour qualifier une très bonne copie ou une mauvaise qui aura une influence sur les suivantes. Ainsi, la note d'une copie sera fonction de la (ou des) copie(s) précédente(s) et leur est le plus souvent opposée. Par exemple, une copie « moyenne » précédée par une copie « excellente » sera certainement jugée comme « faible » alors qu'elle aurait été précédée par une copie « médiocre », elle aurait pu être jugée comme « moyennement forte ». Nous parlons ici, *des effets dits de contraste.*



**Figure 3 : Modélisation des effets de contraste dans la correction d'une série de copie de mathématiques.** Y. Abernot (1996)

*L'effet des « ancrages hautes » (copies 4, 5 et 6) se répercute (par contraste) sur les copies 7 et 8.*

## **1.2 QUESTION DE RECHERCHE ET HYPOTHESES**

Dans la présente recherche, nous nous intéressons à la question de la validité en évaluation et plus particulièrement aux effets d'ordre (ou de position) au sens strict, nous ne chercherons donc pas à mettre en avant les effets de contraste.

En outre, grâce à cette étude, nous souhaiterions savoir si l'influence des effets d'ordre est plus ou moins importante en fonction du type de discipline (disciplines dites « normées » type mathématiques, physique-chimie et disciplines dites « moins normées » type français, philosophie). C'est pourquoi notre question de recherche est la suivante :

**L'effet de position est-il moindre dans des disciplines dites « normées » (type mathématiques ou physique-chimie) par rapport à celui constaté dans d'autres disciplines dites « moins normées » (type français, philosophie) ?**

Pour répondre au mieux à cette question, nous avons émis plusieurs hypothèses :

**Hypothèse 1** : Au fur et à mesure de la correction la moyenne des notes diminue de façon plus significative en français/philosophie qu'en mathématiques/physique-chimie.

**Hypothèse 2** : Au fur et à mesure de la correction la moyenne des notes diminue de la même façon (dans les mêmes proportions) dans les deux types de disciplines.

**Hypothèse 3** : Au fur et à mesure de la correction la moyenne des notes diminue de façon plus significative en mathématiques/physique-chimie qu'en français/philosophie.

## **2. RECUEIL ET TRAITEMENT DES DONNEES DE TERRAIN**

Comme nous l'avons expliqué précédemment, JJ Bonniol est un docimologue reconnu pour son travail sur les « effets d'ordre et d'ancrage ». Ainsi, j'ai souhaité conduire mon étude sur le modèle de celles menées par JJ. Bonniol lors de sa thèse en 1972 mais dans un cadre bien plus restreint ; Car je ne pouvais (pour des raisons d'organisations évidentes) réunir plusieurs groupes d'une dizaine de correcteurs des quatre disciplines étudiées afin de leur faire corriger plusieurs séries de copies dans des ordres inversés.

De ce fait, le protocole expérimental utilisé dans cette recherche consiste à recueillir des données de correction de « terrain », c'est-à-dire des séries de notes attribuées par des enseignants - de mathématiques, physique-chimie, français et philosophie - lors de la correction de piles de copies de classes dont ils ont la charge.

Il est important de souligner que les données recueillies sont des données de « terrain », des notes réelles et que par leurs natures, elles ont sûrement été contaminées à la fois par des effets parasites tel que l'effet d'ordre (effet étudié ici) mais aussi - et même sûrement - par d'autres effets comme l'effet de stéréotypie, l'effet de « halo », (...). De plus, comme précise judicieusement Pierre Merle (2007) les notes sont « fabriquées » car elles sont issues de « bricolage », d'arrangements internes et externes. Ainsi, les résultats que nous obtiendrons à l'issue de cette étude seront à relativiser (à cause des conditions environnementales de recueil différentes).

### **2.1 METHODOLOGIE EMPLOYEE POUR RECUEILLIR LES DONNEES**

Pour recueillir les données expérimentales nécessaires à mon étude, j'ai sollicité des enseignants des quatre disciplines étudiées (tous issus de l'enseignement agricole) via un formulaire (voir annexe 1) dans lequel je leur ai demandé de renseigner, en plus des notes qu'ils ont attribué à chaque copie en fonction de son rang, le nombre d'année d'expérience du correcteur, le mode de correction suivi (copie par copie ou exercice par exercice) et si la correction a été effectuée en une ou plusieurs fois. Ces deux dernières précisions peuvent paraître relever du détail, mais au contraire, elles peuvent avoir de grosses influences sur l'effet d'ordre que l'on cherche à mettre en évidence. *Nous reprendrons cette explication plus en détail dans la partie 2.3 : Méthodologie employée pour analyser les données.*

Le formulaire, accompagné de sa « notice d'utilisation » (voir annexe 2) a été envoyé simultanément, par e-mail, à 30 enseignants de mathématiques, 10 enseignants de physique-chimie, 25 enseignants de français et 15 enseignants de philosophie. De plus, j'ai posté le même message à plusieurs reprises sur les quatre conférences disciplinaires dans l'objectif d'élargir mon échantillon expérimental.

Le période de recueil s'étendait du 21 novembre 2016 au 10 février 2017. Cet intervalle comprenait deux périodes de vacances scolaires (vacances de Noël et de février), période propice à la correction d'évaluation. J'espérais ainsi qu'un maximum d'enseignants puisse participer à cette expérimentation.

A l'issue de la période de recueil, j'ai comptabilisé 21 enseignants participants dont 15 enseignants de mathématiques, un de physique-chimie et cinq de français, qui m'ont retourné respectivement : 43 questionnaires pour les enseignants de mathématiques, 7 questionnaires pour l'enseignant de physique-chimie et 6 questionnaires pour les enseignants de français.

Etant donné la distribution de notre échantillonnage, nous pouvons dès à présent déplorer la faible confiance que nous pourrions accorder aux résultats de cette étude de par le peu de quantité de données recueillies. En effet, pour être représentative, une étude statistique doit porter sur un grand nombre d'échantillons.

Il est important de préciser que les enseignants participants n'ont pas eu connaissance du motif de cette étude. Tout cela, dans l'objectif de ne pas biaiser d'avantage les données recueillies.

## 2.2 METHODOLOGIE EMPLOYEE POUR TRAITER LES DONNEES

Dès réception des questionnaires un code a été attribué à chaque correcteur afin de respecter leur anonymat lors du traitement et de l'exploitation des données.

Ce code personnel comprend une ou plusieurs lettres ainsi qu'un nombre. Par exemple, le premier correcteur de français s'est vu attribué le code F1, celui de physique-chimie le code PC1...

Dans un second temps, il a fallu classer les questionnaires en fonction de leur mode de correction : exercice par exercice ou copie par copie.

Le tableau ci-dessous présente les 56 séries de copies reçues en fonction du type de discipline (normée ou moins normée), du correcteur et du mode de correction suivi.

Correction exercice par exercice	Correction copie par copie
<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M1 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M2 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M4 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M5 : 2</div> </div>	<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M3 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M4 : 7</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M7 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M9 : 1</div> </div>
<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M6 : 2</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M7 : 7</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M8 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M11 : 4</div> </div>	<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M10 : 5</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M11 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M12 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M14 : 1</div> </div>
<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M12 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M13 : 1</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">PC1 : 4</div> </div>	<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">M15 : 4</div> <div style="border: 1px solid gray; padding: 5px; margin: 5px;">PC1 : 3</div> <div style="background-color: green; padding: 5px; margin: 5px;">F1 : 1</div> <div style="background-color: green; padding: 5px; margin: 5px;">F2 : 1</div> </div>
25 séries de copies	31 séries de copies
<i>Total : 56 séries de copies</i>	
Disciplines dites « moins normées »	Disciplines dites « normées »

**Tableau 7 : Origine des séries de copies étudiées en fonction du type de discipline et du mode de correction suivi.**

A la lecture de ce tableau, nous remarquons que certains enseignants des disciplines dites « normées » corrigent différemment, en fonction des séries de copies, soit en suivant le mode « exercice par exercice » ou « copie par copie » à l'instar des enseignants M4, M7, M11, M12 ou PC1. En revanche, les cinq enseignants de français ont toujours corrigé leurs séries de copies en suivant le mode « copie par copie ».

Nous pouvons imaginer que cette différence dans le mode de correction suivi provienne du type d'exercice demandé dans chaque évaluation en fonction du type de disciplines. En effet, si les évaluations de français étudiées ici portaient sur un exercice de dissertation, il semble cohérent que les enseignants aient suivi un mode de correction « copie par copie ». A l'inverse de leurs confrères enseignants de mathématiques ou physique-chimie qui proposent, en règle générale, plusieurs exercices indépendants dans une même évaluation.

Avant d'arriver au traitement final des données, nous avons réalisé un prétraitement de celles-ci. Ce dernier consistait à opérer un « centrage-réduction » des données en ramenant la moyenne de l'ensemble des séries de notes à 10 (centrage) et l'écart-type à 3 (réduction) comme l'avaient réalisé G. Noizet et JP. Caverni (1978) lors du traitement des données recueillies par JJ. Bonniol (1965) (cf. tableau 6 page.21).

Cette opération de centrage-réduction est effectuée afin de rendre « plus comparable » l'ensemble des séries de notes (issues de différents correcteurs, établissements et classes) entre elles, en se libérant notamment de l'influence des comportements de « générosité ou sévérité » des correcteurs ainsi que des différences de niveau qu'il peut y avoir d'une classe à l'autre voire d'un établissement à l'autre.

A la suite de cette étape de « prétraitement », très importante dans notre étude, nous avons mis en place deux méthodes différentes pour traiter nos séries de notes :

- **Méthode 1 : Traitement des données par sous-ensemble.**

Nous avons choisi de conduire cette méthode en priorité car c'est celle qui a été utilisée par G. Noizet et JP. Caverni (1978) dont les résultats sont présentés dans le tableau 6 page. 21.

Pour ce faire, nous avons divisé chaque série de notes en sous-ensembles égaux (ou presque selon si la quantité de copies de la série était un nombre pair ou impair) : soit en deux afin d'obtenir la première moitié et la seconde moitié, soit en quatre pour obtenir premier quart, second quart, troisième quart et quatrième quart.

Ensuite, nous avons calculé la moyenne de chacun des sous-ensembles en fonction du type de discipline afin d'étudier les probables variations de celle-ci compte tenu du sous-ensemble considéré et du type de discipline (« normée » vs « moins normée »).



- **Méthode 2 : Traitement des données par rangs.**

Cette méthode consiste à calculer la moyenne des notes rang par rang en fonction du type de discipline (séparation des disciplines « normées » et des disciplines « moins normées ») et de comparer et/ou analyser les potentielles évolutions de la moyenne intra-disciplinaire et interdisciplinaire.

Dans les faits, nous avons calculé la moyenne de toutes les notes de mathématiques et de physique-chimie (ces deux disciplines constituent les disciplines dites « normées ») des copies de rang 1. Puis dans une autre colonne, nous avons réalisé la même opération pour les copies de rang 1 de français (discipline dite « moins normée »). Cette opération a été répétée pour l'ensemble des rangs de copies dans les deux types de disciplines.

Cette méthode nous permet de voir plus en détail l'évolution de la moyenne en fonction du rang et donc de repérer plus facilement l'influence de l'effet de position qu'avec la méthode 1 (traitement des données par sous-ensembles). En revanche, cette méthode a une limite non négligeable : les résultats des derniers rangs (de la copie 25 à la copie 35 par exemple) sont biaisés car le nombre de copies par série est très différent (les enseignants participants n'ont pas tous le même nombre d'élèves en classe). En effet, une moyenne avec 15 notes sera beaucoup plus représentative qu'une moyenne avec 3 voire 2 notes...

### **2.3 METHODOLOGIE EMPLOYEE POUR ANALYSER LES DONNEES**

Cette étude a pour objectif de mettre en évidence l'influence de l'effet d'ordre lors de la correction d'évaluations dans différents types de disciplines.

Pour mesurer réellement cet « effet d'ordre », il faut que chaque copie soit corrigée d'une traite et même que toutes les copies de la série soient corrigées à la suite les unes des autres c'est à dire sans faire de pause.

Ainsi, par logique nous ne pouvons pas rechercher de façon pertinente un « effet d'ordre » dans les séries de copies ayant été corrigées selon le mode « exercice / exercice ». En effet, en suivant ce mode de correction, la pile de copies est retournée à chaque exercice (la première copie corrigée lors du premier exercice deviendra la dernière lors de la correction du second et ainsi de suite...). De ce fait, toutes constatations du phénomène de diminution de la moyenne au fur et à mesure du processus de correction seraient issues du pur hasard.

Nous devons donc raisonnablement exclure les séries de copies corrigées en suivant le mode de correction « exercice / exercice » de notre analyse.

### 3. ANALYSE ET DISCUSSION DES RESULTATS DE L'ETUDE

Dans cette dernière partie, nous allons nous attacher à analyser le plus clairement possible les données recueillies lors de notre enquête. Pour ce faire, nous devons classer chaque série de copies en fonction de l'allure de sa droite (modélisation graphique de l'évolution de sa moyenne). Ainsi, nous avons séparé les séries de copies en deux : d'un côté celles ayant une moyenne qui diminue au fur et à mesure du processus de correction (droite descendante) et de l'autre, l'ensemble des autres possibilités : droite brisée en dents de scie ou bien moyenne qui a tendance à augmenter au fur et à mesure du processus de correction... A la suite de quoi, nous analyserons la variation de la moyenne entre les sous-ensembles et entre les deux types de disciplines ; tout cela dans l'objectif de répondre au mieux à notre question de recherche.

#### 3.1 ANALYSE SELON LA METHODE 1 : TRAITEMENT DES DONNEES PAR SOUS-ENSEMBLES, SEPARATION DES SERIES EN MOITIÉS.

- **Classification des séries de copies en fonction de l'évolution de leur moyenne au fur et à mesure du processus de correction.**

Vous trouverez en annexe 3 l'ensemble des modélisations graphiques représentant chaque série de copie (selon la méthode 1 : séparation des séries en moitiés), en fonction du type de discipline et du mode de correction suivi. Le tableau ci-dessous a été construit suite à l'observation des modélisations graphiques présentes en annexe 3.

		Séries de copies avec une moyenne ayant tendance à diminuer au fur et à mesure du processus de correction.	Autres cas de figures	Taux de séries de copies ayant une moyenne qui diminue au fur et mesure du processus de correction.
Disciplines dites :	« normées »			<b>52%</b>
	« moins normées »			<b>67%</b>

**Tableau 8 : Distinction des séries de copies présentant un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (probable effet d'ordre). Selon la méthode 1 : séparation des séries en moitiés.**

Grâce au tableau 8, nous pouvons remarquer que seulement 52% des séries de copies issues des disciplines normées analysées (suivant le mode de correction copie par copie) présentent un effet d'ordre probable. C'est-à-dire une diminution de la moyenne au fur et à mesure du processus de correction.

En revanche, pour les disciplines dites « moins normées » ce ratio monte à 67%. Mais, il est important de pondérer ce résultat car notre échantillon pour les disciplines dites « moins normées » est seulement composé de 6 séries de copies. Ainsi, nous pouvons imaginer que ce ratio aurait tendance à se rapprocher des 50% si nous augmentions le nombre de séries de copies.

- **Variation de la moyenne entre les deux sous-ensembles (première et seconde moitié) en fonction du type de discipline**

Afin de confirmer ou d'infirmer une ou l'ensemble de nos trois hypothèses, nous devons mesurer la proportion dans laquelle diminue la moyenne pour chaque type de disciplines au fur et à mesure de la correction.

Pour ce faire, nous allons calculer la variation entre la moyenne de la première moitié de la série et celle de la seconde pour chaque série de copies présentées dans la colonne de gauche du tableau 8 (Distinction des séries de copies présentant un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (probable effet d'ordre). *Selon la méthode 1 : séparation des séries en moitiés.*)

	DISCIPLINES DITES "NORMEES"													DISCIPLINES DITES "MOINS NORMEES"			
Série de copies	M4 (1)	M4 (4)	M4 (5)	M4 (7)	M7	M9	M10 (2)	M10 (3)	M10 (4)	M10 (5)	M15 (2)	PC1 (1)	PC1 (3)	F3	F4 (1)	F4 (2)	F5
Variation de la moyenne: (moyenne seconde moitié - moyenne première moitié)	-3,00	-2,87	-1,27	-0,02	-1,18	-0,67	-1,46	-1,80	-0,08	-0,09	-0,44	-0,21	-1,95	-1,22	-0,44	-1,07	-0,73
Variation moyenne selon le type de disciplines	<b>-1,16</b>													<b>-0,87</b>			

**Tableau 9 : Variation de la moyenne entre le premier et le second sous-ensemble selon le type de discipline.**

Nous pouvons constater que la moyenne des notes des séries de copies issues des disciplines dites « normées » diminue, en moyenne, d'1,16 point entre la première et la seconde moitié de la série, alors qu'elle diminue seulement de 0,87 point pour les disciplines dites « moins normées ».

**Au vu de ces résultats, selon la méthode d'analyse 1** (traitement des données par sous-ensemble : moitiés) **nous pourrions valider l'hypothèse 3** : *Au fur et à mesure de la correction la moyenne des notes diminue de façon plus significative en mathématiques/physique-chimie (disciplines dites « normées ») qu'en français/philosophie (disciplines dites « moins normées »).*

Or, il est important de souligner que ces résultats portent sur un échantillon de départ limité et sur une faible part de cet échantillon. Pour rappel, seulement 52% des 50 séries de copies issues des disciplines dites « normées » présentaient un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (c'est-à-dire un potentiel effet d'ordre).

D'ailleurs, afin de savoir si le phénomène de diminution de la moyenne constaté dans les séries de copies corrigées suivant le mode « copie par copie » est la résultante certaine de l'effet d'ordre ou de position décrit par les docimologues, j'ai décidé de mesurer le taux de séries de copies ayant une moyenne qui diminue au fur et mesure du processus de correction dans l'ensemble des séries de copies corrigées selon le mode « exercice par exercice ». *Pour rappel, nous avons précédemment exclu ces séries de copies de notre analyse après avoir démontré qu'en suivant ce mode de correction, tout phénomène de diminution de la moyenne au fur et à mesure de la correction est le fruit du hasard.*

A ma grande surprise, le résultat est le même : 13 séries de copies sur 25 (soit 52%) présentent une diminution de la moyenne entre le premier et le second sous-ensemble.

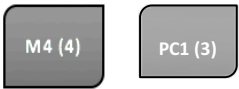
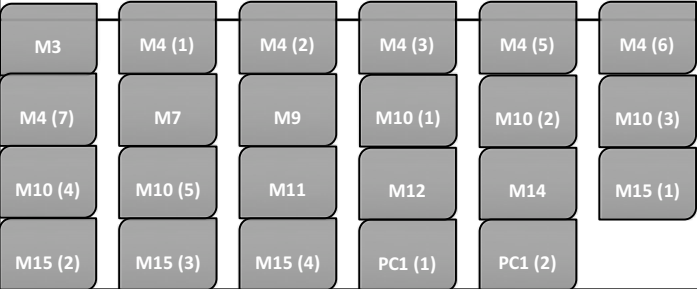
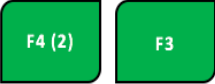

Au regard de ce nouveau constat, nous ne pouvons pas affirmer que le phénomène de diminution de la moyenne constaté entre le premier et le second sous-ensemble dans les séries de copies corrigées selon le mode « copie par copie » soit directement lié à l'effet de position (ou effet d'ordre) ; même si nous ne pouvons pas l'infirmier avec certitude non plus.

**Ainsi, d'après la méthode d'analyse 1** (traitement des données par sous-ensemble : séparation des séries en moitiés) **aucune de nos trois hypothèses de départ ne peut être confirmée.**

### 3.2 ANALYSE SELON LA METHODE 1 : TRAITEMENT DES DONNEES PAR SOUS-ENSEMBLE, SEPARATION DES SERIES EN QUARTS.

- Classification des séries de copies en fonction de l'évolution de leur moyenne au fur et à mesure du processus de correction.

*Vous trouverez en annexe 4 l'ensemble des modélisations graphiques représentant chaque série de copie (selon la méthode 1 : séparation des séries en quarts), en fonction du type de discipline et du mode de correction suivi. Le tableau 10 (ci-dessous) a été construit suite à l'observation des modélisations graphiques présentes en annexe 4.*

Séries de copies avec une moyenne ayant tendance à diminuer au fur et à mesure du processus de correction.	Autres cas de figure (tendance à la hausse, droite brisée : dents de scie...)	Taux de séries de copies ayant une moyenne qui diminue au fur et mesure du processus de correction.
		8 %
		33,3 %

**Tableau 10 : Distinction des séries de copies présentant un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (probable effet d'ordre). Selon la méthode 1 : séparation des séries en quarts.**

Les résultats présentés dans ce tableau ont de quoi nous laisser perplexe : seulement 8% (soit 2/25) des séries de copies issues des disciplines dites « normées » et corrigées selon le mode « copie par copie » présentent une droite descendante lors de leurs modélisations graphiques. Ce ratio passe à 33,3% (soit 2/6) pour les séries de copies issues des disciplines dites « moins normées ». Malgré tout, ces résultats sont cohérents. En effet, nous avons divisé par 4 l'ensemble des séries de copies sans tenir compte de leurs effectifs respectifs. Malheureusement, dans notre étude nous avons des séries de copies avec des effectifs très différents et même certaines avec moins de 15 copies.

Ainsi, effectuer un classement seulement au regard des modélisations graphiques (comme c'est le cas dans le tableau 10) me paraît peut pertinent.

C'est pourquoi, j'ai décidé de me fier à la variation de la moyenne de façon globale entre les 4 quarts.

- **Variation de la moyenne entre les quatre sous-ensembles (premier, second, troisième et quatrième quart) en fonction du type de discipline :**

La variation est un indicateur qui permet de mesurer l'évolution d'une variable par rapport à sa valeur de départ. La formule est la suivante : valeur d'arrivée – valeur de départ. Or dans notre cas, nous avons quatre valeurs. Nous allons donc réaliser la moyenne de la variation de chaque quart entre eux. Cela se traduit ainsi : ((moy. 4ème quart - moy. 3ème quart) + (moy. 3ème quart - moy. 2nd quart) + (moy. 2nd quart - moy. 1er quart)) / 3.

En suivant cette démarche, nous nous sommes rendu compte qu'au fur et à mesure du processus de correction, même en séparant les séries en quatre quarts, il y a un phénomène de diminution de la moyenne mais ce dernier reste faible : -0,59 point en moyenne pour les disciplines dites « normées » et – 0,36 point en moyenne pour les disciplines dites « moins normées ».

L'ensemble de ces résultats sont regroupés dans le tableau 11 ci-dessous.

	DISCIPLINES DITES "NORMÉES"												DISCIPLINES DITES "MOINS NORMÉES"				
Série de copies	M4 (1)	M4 (3)	M4 (4)	M4 (5)	M4 (6)	M7	M9	M10 (1)	M10 (2)	M10 (3)	M10 (5)	PC1 (3)	F2	F3	F4 (1)	F4 (2)	F5
<b>Variation de la moyenne:</b> ((moy. 4ème quart - moy. 3ème quart)+(moy. 3ème quart - moy. 2nd quart)+(moy. 2nd quart - moy. 1er quart)) / 3	-0,37	-0,25	-2,02	-0,48	-0,33	-0,48	-0,44	-0,04	-0,53	-0,45	-0,43	-1,28	-0,17	-0,72	-0,26	-0,48	-0,19
<b>Variation moyenne selon le type de disciplines</b>	<b>-0,59</b>												<b>-0,36</b>				

**Tableau 11 : Variation de la moyenne entre le premier et le quatrième sous-ensemble selon le type de discipline.**

Notons qu'en suivant la démarche « mesure de la variation de la moyenne », plutôt que celle de la modélisation graphique, nous obtenons 12 / 25 (soit 50%) séries de copies présentant une diminution de la moyenne au fur et à mesure du processus de correction pour les disciplines dites « normées » et 5/6 (soit 83,3%) pour les disciplines dites « moins normées » en suivant bien sûr, la méthode d'analyse 1 : séparation des séries de données en quarts. Ces résultats s'apparentent à ceux obtenus lorsque nous avons suivis la méthode d'analyse 1 : séparation des séries de données en moitiés : 52% (soit 13/25) pour les disciplines dites « normées » et 67% (soit 4/6) pour les disciplines « moins normées ».

D'ailleurs, si nous nous observons seulement la dernière ligne du tableau 11 (Variation moyenne de la moyenne au fur et à mesure du processus de correction selon le type de discipline), **nous pourrions encore valider l'hypothèse 3** : *Au fur et à mesure de la correction la moyenne des notes diminue de façon plus significative en mathématiques/physique-chimie (disciplines dites « normées ») qu'en français/philosophie (disciplines dites « moins normées »)*. Car nous constatons que dans les disciplines dites « normées » la variation moyenne entre les quatre sous-ensembles est plus importante (-0,59 point) que celle constatée dans les disciplines dites « moins normées » (-0,36 point).

### **3.3 ANALYSE SELON LA METHODE 2 : TRAITEMENT DES DONNEES PAR RANG**

Cette méthode d'analyse est très différente des deux précédentes. En effet, cette dernière consiste à calculer la moyenne des notes rang par rang, c'est-à-dire toutes séries de notes confondues (en respectant tout de même le type de discipline).

Comme nous l'avons souligné dans la partie 2.2 (*Méthodologie employée pour traiter les données*), cette méthode d'analyse se heurte à une limite non négligeable : l'hétérogénéité des effectifs par séries.

L'effet d'ordre (ou de position) repose sur le postulat suivant : *Les correcteurs seraient plus sévères en fin de série qu'au début*. Or, nous ne pouvons pas espérer observer un tel phénomène si nous n'étudions pas des séries de notes comportant approximativement le même nombre de copies (+ ou – 5 copies d'écart par exemple). En effet, pour une série de 12 copies, la 12<sup>ème</sup> et dernière copie souffrira peut-être d'un effet d'ordre mais ce dernier n'apparaîtra pas car il sera noyé dans d'autres séries de copies pour lesquelles la 12<sup>ème</sup> copie correspond seulement au milieu de la série (moins sujette aux effets d'ordre).

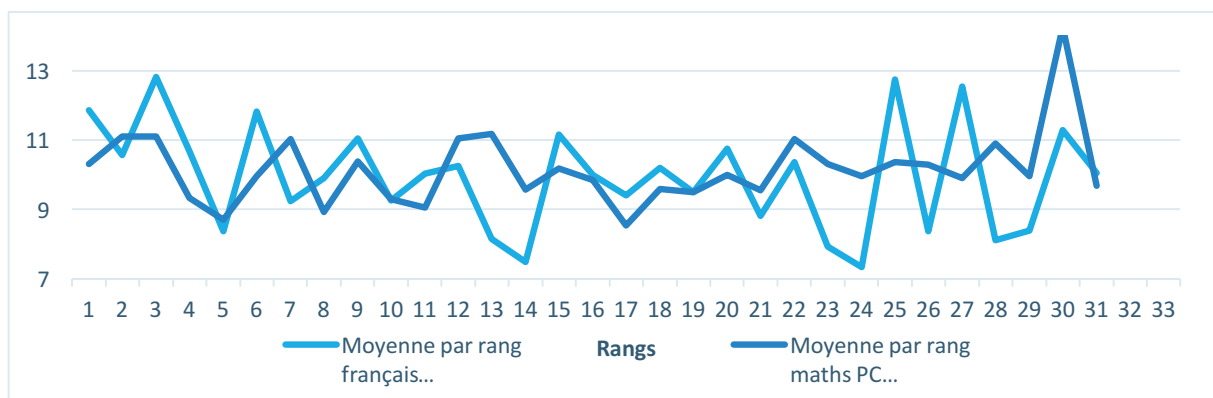
Ainsi, dans l'objectif de limiter les biais liés à l'hétérogénéité des effectifs par série, nous allons prendre en compte seulement les séries ayant un effectif compris entre 20 et 32 copies. Il est vrai que cette amplitude dépasse largement nos préconisations précédentes de + ou – 5 copies mais si nous nous en tenons à cette faible amplitude nous aurions trop peu de séries pour réaliser notre analyse.

*Vous trouverez tout de même, en annexe 5, les graphiques présentant l'évolution de la moyenne par rang au fur et à mesure du processus de correction en fonction du type de discipline et du*

*mode de correction suivi. Ces graphiques prennent en compte l'ensemble des séries de notes quel que soit leur effectif.*

Après avoir retiré les séries ayant un faible effectif, nous nous retrouvons avec 15 séries (soit 60% des séries de copies) pour les disciplines dites « normées » et 5 séries (soit 83,3%) pour les disciplines dites « moins normées ».

Le graphique obtenu en suivant la 2<sup>nde</sup> méthode de traitement des données est présenté dans la figure 4 ci-dessous.



**Figure 4 : Evolution de la moyenne des notes par rang au fur et à mesure du processus de correction en fonction du type de discipline.**

A la lecture de ce dernier, il nous est très difficile de réaliser une analyse claire qui nous permettrait de valider une de nos trois hypothèses de départ. En effet, les courbes présentées ne laissent pas apparaître de tendance franche (soit à la hausse, soit à la baisse). Au contraire, elles sont toutes les deux en dents de scie et très proche l'une de l'autre.

Ceci étant, les correcteurs des disciplines dites « moins normées » semblent être plus soumis aux effets de contraste car l'amplitude de la note moyenne entre deux rangs est plus importante que celle constatée dans les disciplines dites « normées ». Malgré tout, ce constat doit être relativisé car nous sommes sur un très faible échantillon pour les disciplines dites « moins normées » (5 séries seulement) donc ce dernier est très peu représentatif.

Contrairement aux méthodes d'analyse précédentes, celle-ci ne nous a pas permis d'identifier une hypothèse comme potentiellement probable. Pour être pertinente et efficace, je pense qu'il faudrait appliquer cette méthode d'analyse seulement lorsque nous avons un large échantillon (>50 séries minimum) et que l'ensemble des séries de copies soient approximativement de taille similaire.



## PERSPECTIVES DES SUITES DE RECHERCHE

Le présent travail n'est qu'une amorce, une introduction à un travail de recherche approfondi. En effet, pour parler de faits observés de façon « scientifique » et d'en tirer des conclusions pertinentes, il faudrait mener une étude bien plus poussée que celle-ci.

Tout d'abord, ce serait un travail de longue haleine, bien supérieur aux six mois dont nous disposons pour réaliser cette étude, cela pourrait donc tout à fait être réalisé dans le cadre d'une thèse par exemple.

De plus, c'est le protocole expérimental de recueil des données dans sa globalité qui est à revoir. En effet, pour pouvoir réaliser des comparaisons et des analyses avisées, il faut absolument utiliser des données semblables et « comparables » (et non des données de terrain comme ce fut le cas dans la présente étude). Dans les faits, il faudrait réunir plusieurs groupes de correcteurs (de taille conséquente afin d'obtenir des résultats représentatifs) des deux types de disciplines et leur soumettre des séries de copies identiques dans des ordres de corrections différents. Ceci étant la méthode de recueil de données mise en place par JJ. Bonniol (1972).

En outre, pour observer un effet parasite bien particulier (ici, l'effet d'ordre ou de position), il faut limiter au maximum les risques de contaminations par d'autres effets. Utiliser le protocole de JJ. Bonniol participe à limiter ces risques (environnement identique, copies anonymes...).

## LIMITES ET DIFFICULTES RENCONTREES DURANT CE TRAVAIL

**-Le temps.** La variable temps a été pour moi la plus grande limite à laquelle je me suis confrontée durant la réalisation de ce travail. En effet, c'est très frustrant de devoir limiter ses lectures et ses recherches bibliographiques à cause du temps qui file. De plus, je n'ai disposé que de quelques semaines pour recueillir les données expérimentales (afin d'avoir assez de temps pour pouvoir les analyser), ce qui nous conduit tout naturellement à la seconde limite.

**- La faible quantité de données expérimentales.** Malgré plusieurs envois (mail nominatif, mail commun sur les conférences disciplinaires), j'ai obtenu très peu de données issues des disciplines dites « moins normées ». En effet, seulement 6 séries de notes m'ont été retournées pour les disciplines dites « moins normées » contre 50 séries de notes pour les disciplines dites « normées ». Il est donc très compliqué de réaliser une analyse pertinente avec si peu de données.

**-L'origine des données expérimentales.** Rappelons que les données traitées dans cette étude sont des données de « terrain ». Ainsi, de par leur nature, elles sont difficilement comparables et surtout elles ont, très certainement, été contaminées par un autre effet parasite que celui étudié (comme l'effet de halo, de stéréotypie par exemple)

## CONCLUSION : PERSPECTIVES D'EXPLOITATION PROFESSIONNELLE

Dans la présente recherche, les effets d'ordre au sens stricts tels que décrits par JJ. Bonniol (1972) et leurs relativités en fonction du type de discipline n'ont pu être mis clairement en évidence, même si malgré tout, la méthode d'analyse 1 (traitement des données par sous-ensemble) laisse apparaître comme potentiellement probable l'hypothèse 3 selon laquelle : *Au fur et à mesure de la correction la moyenne des notes diminue de façon plus significative en mathématiques/physique-chimie (disciplines dites « normées ») qu'en français/philosophie (disciplines dites « moins normées »).*

Si cette étude ne permet pas de valider une des hypothèses de départ, elle ne permet pas non plus de les invalider de façon certaine. Nous savions dès le début que nous n'obtiendrons pas de résultats probants lors de cette étude à cause de la nature des données recueillies (données de « terrain »). Malgré tout, cela n'a pas été un vain travail. En effet, comme nous l'avons souligné précédemment, notre question de recherche pourrait tout à fait être reprise lors d'une prochaine étude. Mais pour obtenir des résultats représentatifs il faudra alors adopter la méthodologie de recueil des données suivi par JJ Bonniol (1972) et non la mienne, afin de limiter les risques de contaminations des données par d'autres biais de correction (type effet de halo, de stéréotypie...) et d'obtenir des séries de copies plus homogène ce qui rendra bien plus pertinente l'analyse qui en découlera.

Ce travail de mémoire a été pour moi l'occasion de me poser les vraies questions concernant l'évaluation : sur son rôle auprès des élèves, sur le sens que je souhaite lui donner et la place que je compte lui accorder dans mon enseignement...

Même si je n'ai pas réussi à démontrer clairement la présence des effets d'ordre tout au long de cette étude, j'ai conscience de leur existence tout comme celle des autres types de biais présentés précédemment. C'est pourquoi j'essaie d'adopter une ligne de conduite « particulière » lors de la correction d'évaluation (en respectant les quelques règles suivantes) :

- Parcourir l'ensemble de la série de copies afin de se faire une première idée des réponses apportées par les apprenants et de limiter ainsi le phénomène de sévérité lors de la correction de la première copie.

Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

- Ne pas regarder le nom de l'apprenant afin de limiter les effets de stéréotypie ou de halo.
- Suivre un mode de correction « exercice par exercice » (lorsque cela est possible) afin de limiter l'effet d'ordre ou de position.
- Faire des pauses lorsque cela est nécessaire afin de limiter l'effet d'énerverment ou de fatigue.

En suivant ces quelques règles, j'espère réaliser des évaluations et des corrections plus justes, même si je sais que, quoi que l'on fasse, notre jugement sera toujours altéré par des biais ou des effets particuliers car nous sommes soumis à un environnement très différent en fonction des situations et surtout, nous évoluons avec le temps...

« D'une manière générale, la note comporte toujours une auto-évaluation de l'enseignant. Ainsi l'enseignant consciencieux ne peut que s'interroger sur l'adaptation de ses méthodes et du contenu de son enseignement lorsque la majorité des notes est systématiquement inférieure à la moyenne. Il ne sera pas le seul à s'interroger, car les notes sont aussi un indice de ce qui se passe dans la classe et donne lieu à une évaluation extérieure. Les parents, l'administration s'inquiéteront de notes systématiquement trop hautes ou trop basses. L'image que l'enseignant veut se donner et donner aux autres est donc une variable parasite, sans rapport avec la qualité des productions des élèves, mais qui institue la distribution normale des notes comme règle : peu de très mauvaises notes, beaucoup de moyennes et peu d'excellentes. L'enseignant qui s'en éloigne risque de se faire une réputation parfois injuste et toujours difficile à vivre. » Y. Abernot (1996)

## BIBLIOGRAPHIE

Abernot, Y . (1996). *Les Méthodes d'évaluation scolaire*. Paris : Dunod.

Agazzi, A. (1967). *Les aspects pédagogiques des examens*. Strasbourg : Conseil de l'Europe.

Bonboir, A. (1972). *La docimologie*. Paris : Presse Universitaires de France.

Bonniol J.-J. (1965). Les divergences de notation tenant aux effets d'ordre de la correction ». *Cahiers de psychologie*, N°8, pp.181-188.

Bonniol, J.-J. (1972). *Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires. Etude de quelques-uns de leurs déterminants*. Aix-en-Provence : Université de Provence, thèse de doctorat.

Caverni, J.-P. et al. (1975). *Dépendance des évaluations scolaires par rapport à des évaluations antérieures*. *Le travail humain*, n°38, pp. 213-222.

De Ketele, J.-M. (1982). *Docimologie : Introduction aux concepts et aux pratiques*. Louvain-La-Neuve : Gabey.

De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, n° 103, p. 59-80.

De Landsherre, G. (1980). *Evaluation continue et examens : Précis de docimologie*. Bruxelles-Paris : Labor-Nathan.

De Landsheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses Universitaires de France, 2ème édition revue et augmentée.

Hadji, C. (1989). *L'évaluation règles du jeu des intentions aux outils*. Paris : ESF éditeur.

Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* Paris-Bruxelles : De Boeck.

Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.

Merle, P. (2007). *Les notes. Secrets de fabrications*. Paris : Presses Universitaires de France.

Merle, P. (2012). L'évaluation par les notes : quelle fiabilité et quelles réformes ?. *Regards croisés sur l'économie*, vol. 2, n° 12, p. 218-230.

Musial, M, Pradère, F, Tricot, A. (2012). *Comment concevoir un enseignement ?* Bruxelles : De Boeck Supérieur.

Noizet, G, Bonniol, JJ. (1968-1969). Pour une docimologie expérimentale, in *Bulletin de Psychologie*, p. 782-787.

Noizet, G, Caverni, JP. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.

Pieron, H. (1963). *Examens et docimologie*. Paris : Presses Universitaires de France.

Rey, O, Feyfant, A. (2014). Évaluer pour (mieux) former. *Dossier de veille de l'IFÉ*, n°94, septembre. Lyon : ENS de Lyon.

## SITOGRAPHIE

Coron, Mireille. Le modèle de référence professionnel : Déterminant de l'évaluation sommative dans l'enseignement infirmier. <http://fulltext.bdsp.ehesp.fr/Rsi/20/11.pdf>. Consulté le 12/01/2017.

Ministère de l'Éducation Nationale. Extraits du code de l'éducation. <https://www.legifrance.gouv.fr>.

## **TABLE DES FIGURES**

Figure 1 : La double articulation dans l'opération d'évaluation.

Figure 2 : Modélisation des effets d'ordre au sens strict dans le processus de correction d'une série de copies.

Figure 3 : Modélisation des effets de contraste dans la correction d'une série de copies de mathématiques.

Figure 4 : Evolution de la moyenne des notes par rang au fur et à mesure du processus de correction en fonction du type de discipline.

## **TABLE DES TABLEAUX**

Tableau 1 : Fonctions, types et démarches de l'évaluation selon sa place par rapport à la séquence de formation.

Tableau 2 : Ecart moyen, maximums et les plus fréquemment relevés entre les cinq correcteurs.

Tableau 3 : Résultats des différents candidats aux épreuves de français, philosophie, mathématiques, physique-chimie du baccalauréat en fonction des correcteurs.

Tableau 4 : Nombre de correcteurs nécessaires pour obtenir une « note vraie » ou stabiliser la note.

Tableau 5 : Ecart-types observés sur des séries de notes relevées lors de l'enquête Carnegie.

Tableau 6 : Moyenne des notes attribuées par les 18 correcteurs, selon l'ordre de correction dans le cadre de l'étude de JJ. Bonniol (1965).

Tableau 7 : Origine des séries de copies étudiées en fonction du type de discipline et du mode de correction suivi.

Tableau 8 : Distinction des séries de copies présentant un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (probable effet d'ordre). *Selon la méthode 1 : séparation des séries en moitiés.*

Tableau 9 : Variation de la moyenne entre le premier et le second sous-ensemble selon le type de discipline.

Tableau 10 : Distinction des séries de copies présentant un phénomène de diminution de la moyenne au fur et à mesure du processus de correction (probable effet d'ordre). *Selon la méthode 1 : séparation des séries en quarts.*

Tableau 11 : Variation de la moyenne entre le premier et le quatrième sous-ensemble selon le type de discipline.



## TABLE DES MATIERES

<b>INTRODUCTION</b> .....	<b>2</b>
<b>1. CADRE THEORIQUE ET QUESTIONNEMENT</b> .....	<b>3</b>
1.1 CONTEXTE .....	3
1.1.1 <i>L'évaluation, pilier du système éducatif français</i> : .....	3
1.1.2 <i>La docimologie et ses apports sur la question de l'évaluation</i> .....	11
1.2 QUESTION DE RECHERCHE ET HYPOTHESES .....	19
<b>2. RECUEIL ET TRAITEMENT DES DONNEES DE TERRAIN</b> .....	<b>20</b>
2.1 METHODOLOGIE EMPLOYEE POUR RECUEILLIR LES DONNEES .....	20
2.2 METHODOLOGIE EMPLOYEE POUR TRAITER LES DONNEES .....	22
2.3 METHODOLOGIE EMPLOYEE POUR ANALYSER LES DONNEES .....	24
<b>3. ANALYSE ET DISCUSSION DES RESULTATS DE L'ETUDE</b> .....	<b>25</b>
3.1 ANALYSE SELON LA METHODE 1 : TRAITEMENT DES DONNEES PAR SOUS-ENSEMBLES, SEPARATION DES SERIES EN MOITIES. ....	25
3.2 ANALYSE SELON LA METHODE 1 : TRAITEMENT DES DONNEES PAR SOUS-ENSEMBLE, SEPARATION DES SERIES EN QUARTS.....	28
3.3 ANALYSE SELON LA METHODE 2 : TRAITEMENT DES DONNEES PAR RANG.....	30
<b>PERSPECTIVES DES SUITES DE RECHERCHE</b> .....	<b>32</b>
<b>LIMITES ET DIFFICULTES RENCONTREES DURANT CE TRAVAIL</b> .....	<b>33</b>
<b>CONCLUSION : PERSPECTIVES D'EXPLOITATION PROFESSIONNELLE</b> .....	<b>34</b>
<b>BIBLIOGRAPHIE</b> .....	<b>36</b>
<b>SITOGRAPHIE</b> .....	<b>37</b>
<b>TABLE DES FIGURES</b> .....	<b>38</b>
<b>TABLE DES TABLEAUX</b> .....	<b>38</b>
<b>TABLE DES MATIERES</b> .....	<b>40</b>
<b>ANNEXES</b>	

## **TABLE DES ANNEXES :**

<b>Annexe 1 : Formulaire de recueil des données de « terrain ».....</b>	<b>2</b>
<b>Annexe 2 : Notice d'utilisation jointe au formulaire de recueil. ....</b>	<b>3</b>
<b>Annexe 3 : Modélisations graphiques des données recueillies par séries selon la méthode 1 : <i>traitement des données par sous-ensembles, séparation des séries en moitiés</i>.....</b>	<b>4</b>
<b><u>Annexe 4</u> : Modélisations graphiques des données recueillies par séries selon la méthode 1 : <i>traitement des données par sous-ensembles, séparation des séries en quarts</i>.....</b>	<b>11</b>
<b><u>Annexe 5</u> : Modélisations graphiques des données recueillies par types de disciplines et mode de correction suivi selon la méthode 2 : <i>traitement des données par rangs</i>.....</b>	<b>18</b>

## Annexe 1 : Formulaire de recueil des données de « terrain ».

Projet de mémoire

Amélie Combelles

### FICHE DE COLLECTE DE DONNEES

**Date de correction :**

**Discipline concernée :**  Français

Mathématiques

Philosophie

Physique / Chimie

**Nombre d'années d'expérience :**

0 à 5 ans

5 à 10 ans

10 à 15 ans

+ de 15 ans

**Le devoir a-t'il été corrigé en une seule fois ?**  Oui

Non

Oui mais, j'ai modifié certaines notes ultérieurement.

**Méthode de correction utilisée :**

Copie par copie

Exercice par exercice

Rang de la copie (rang de correction)	Note attribuée /20
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	

## Annexe 2 : Notice d'utilisation jointe au formulaire de recueil.

Projet de mémoire

Amélie Combelles

### NOTICE D'UTILISATION DE LA FICHE DE COLLECTE

---

La fiche de collecte doit être complétée lors de la correction d'une pile de copie d'une classe (devoir sur table ou devoir maison noté sur 20 points).

- **Indiquer la date** de correction de votre pile de copies.
- **Cocher la discipline** concernée et votre nombre d'années d'expérience.
- **Indiquer si le devoir a été corrigé d'une seule traite en cochant la case « oui », sinon, cocher la case « non ».**  
En revanche, si vous avez corrigé le devoir d'une seule traite, mais que vous avez été **amené à modifier certaines notes** (pour diverses raisons : modifications du barème, erreur de correction, réajustement...) ultérieurement à la correction (ou pendant la correction des autres copies) **cocher la case « oui mais, j'ai modifié certaines notes ultérieurement ».**
- **Indiquer la méthode de correction utilisée** en cochant la case correspondante (exercice par exercice ou copie par copie).
- **Compléter le tableau** en reportant les notes attribuées à l'évaluation que vous êtes en train de corriger . **Il est très important de respecter l'ordre des copies** (1<sup>ère</sup> copie corrigée = rang 1, 2<sup>ème</sup> copie corrigée = rang 2...) et de **ne surtout pas trier les notes par ordre croissant ou décroissant ou les copies par ordre alphabétique.**

**Il faut traiter la pile de copie telle que recueillie dans la classe** (ne pas modifier l'ordre des copies entre le ramassage en classe et le remplissage du tableau après la correction).

*Vous pouvez me retourner les fiches de recueil par mail à l'adresse suivante : [amelie.combelle@educagri.fr](mailto:amelie.combelle@educagri.fr).*

Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

### Annexe 3 : Modélisations graphiques des données recueillies par séries selon la méthode 1 : *traitement des données par sous-ensembles, séparation des séries en moitiés.*

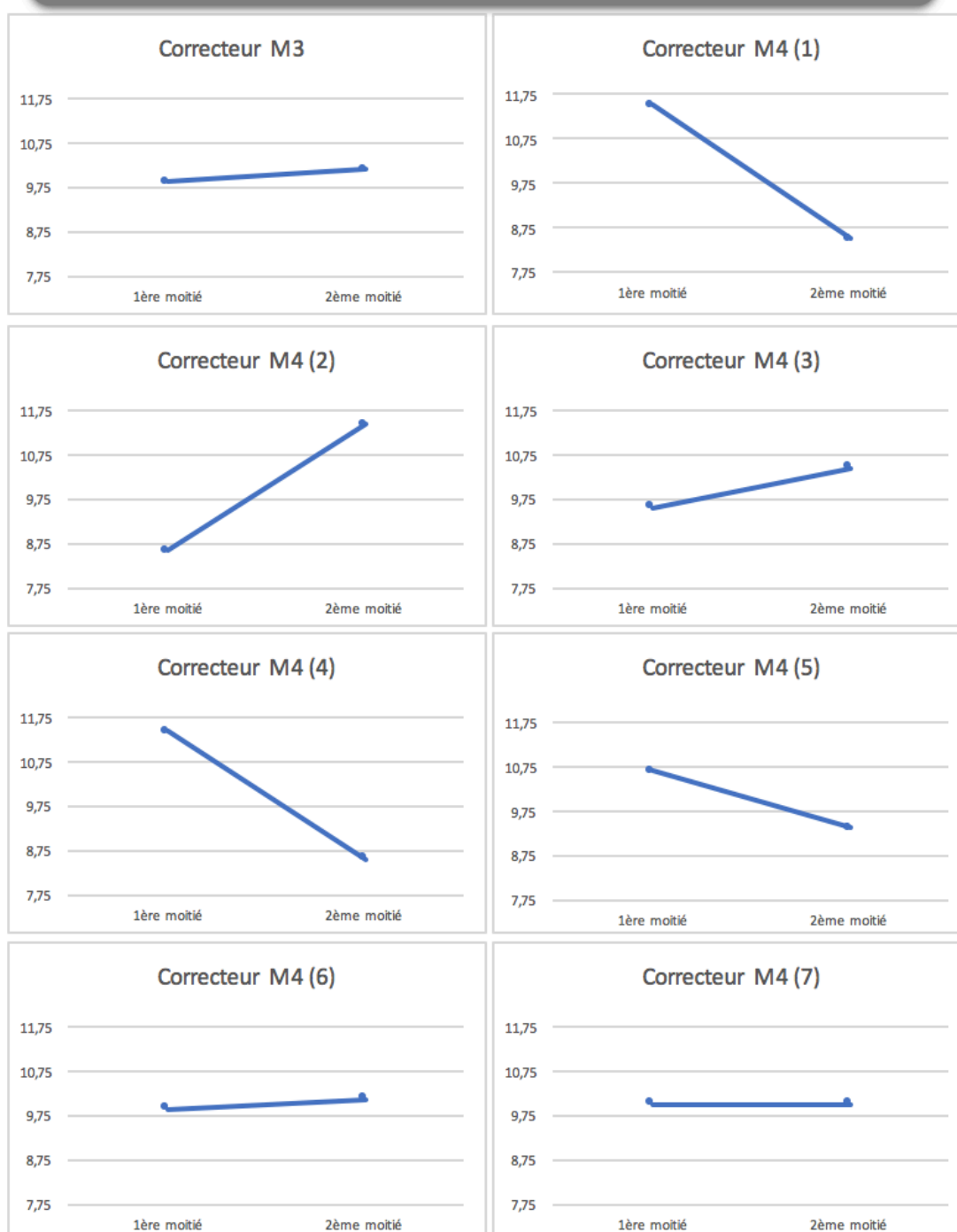
#### Indications de lecture :

La seule droite présente sur chaque graphique représente la moyenne de la série de copie.

Droite bleue = série de copie corrigée en suivant le mode « copie par copie »

Droite jaune = série de copie corrigée en suivant le mode « exercice par exercice »

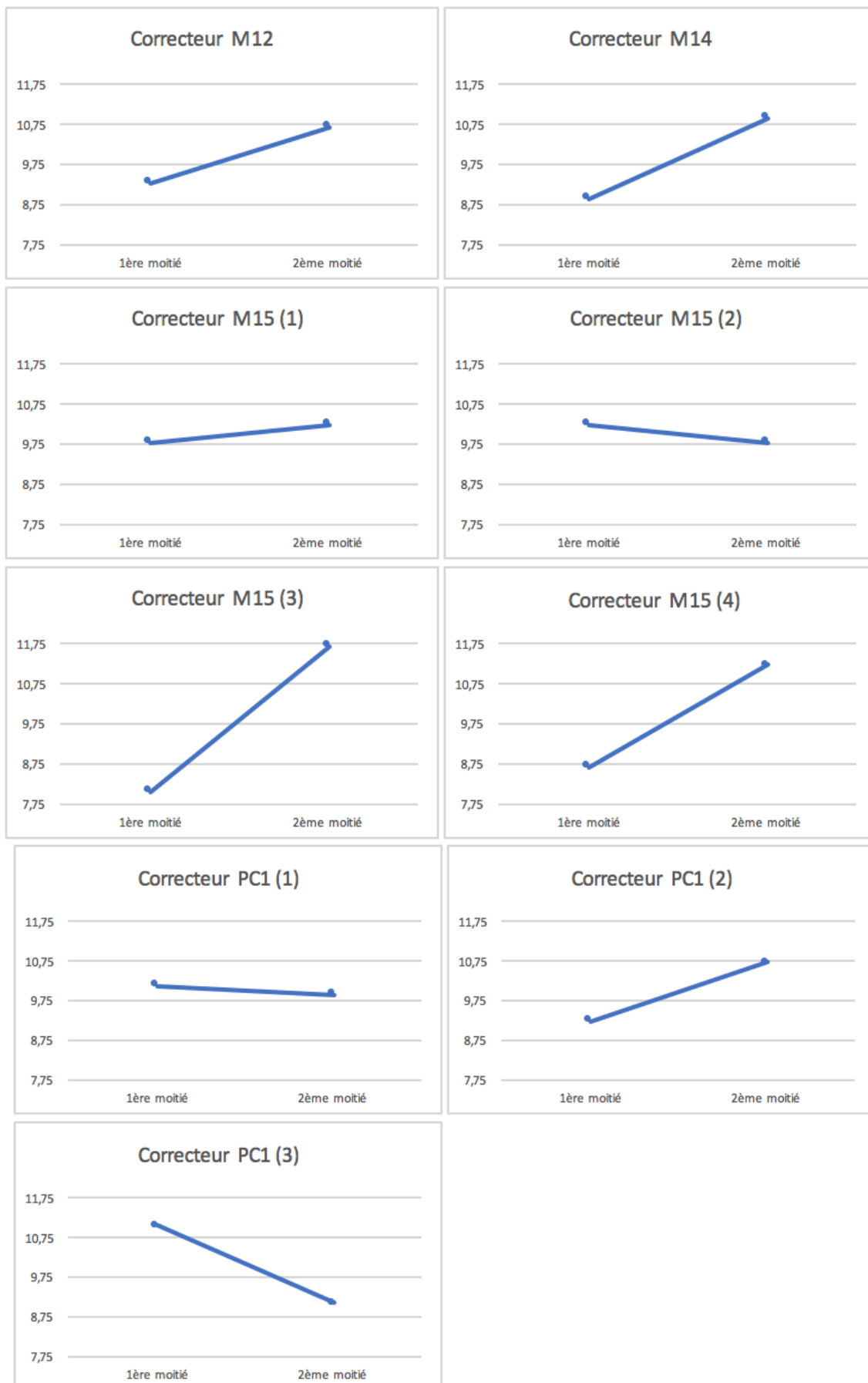
Méthode 1 : Séparation des séries en moitiés.  
Mode de correction : Copie par copie  
Type de disciplines : Disciplines dites « normées »



Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».



Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

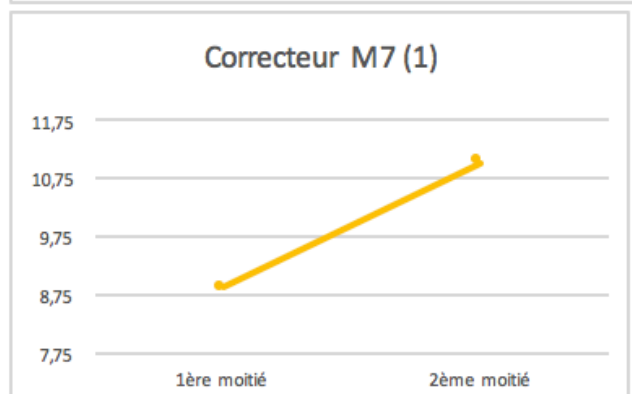
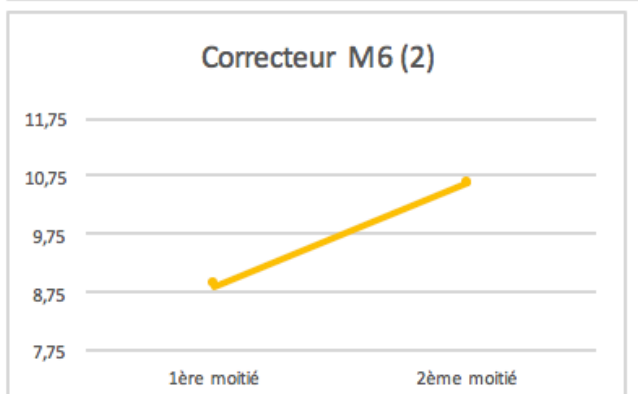
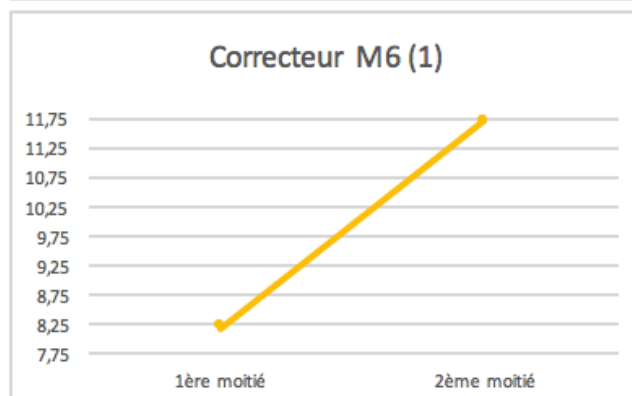
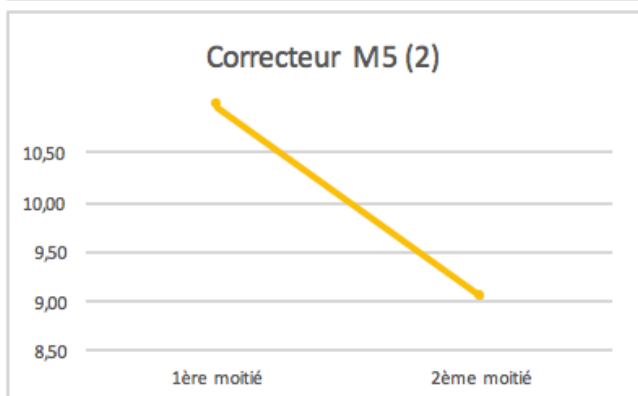
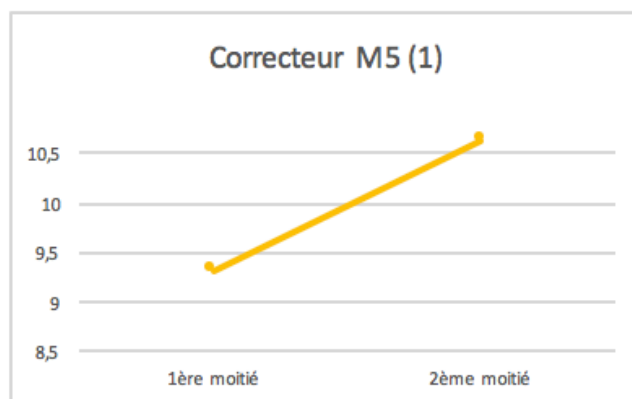
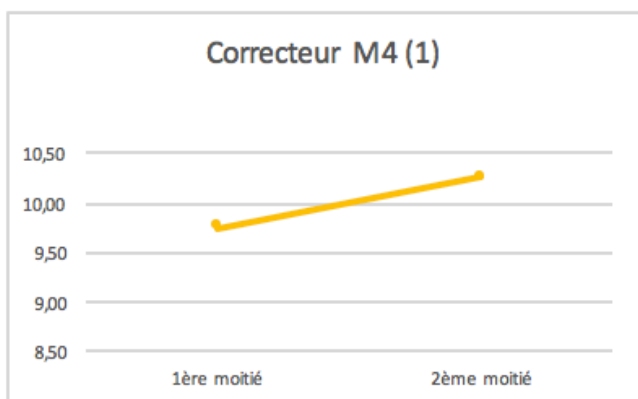
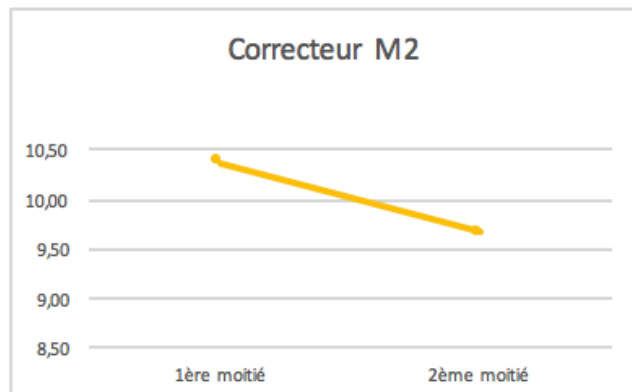
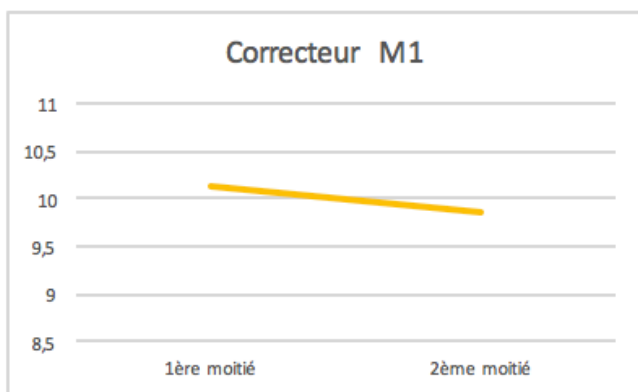


Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

**Méthode 1 : Séparation des séries en moitiés.**

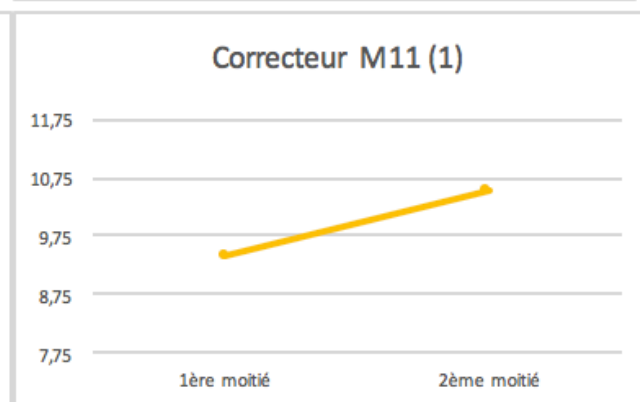
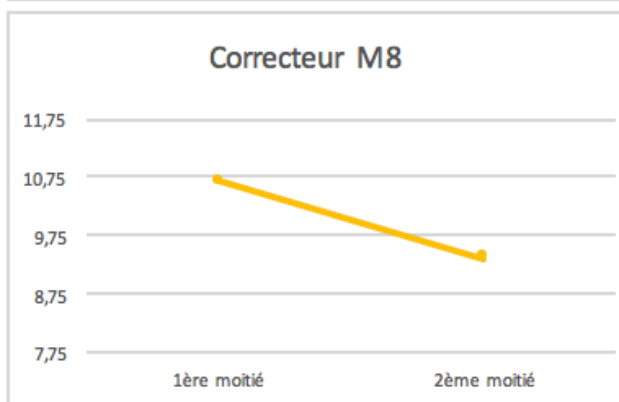
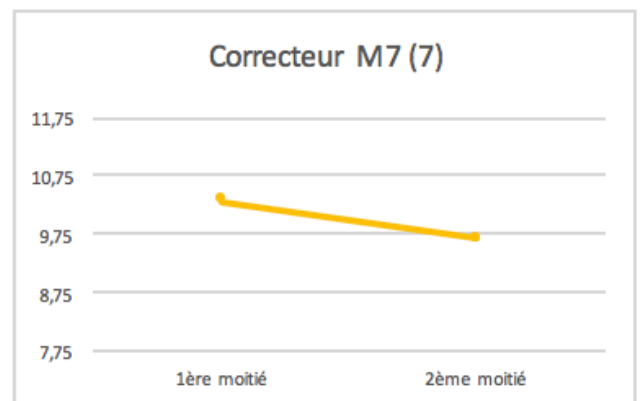
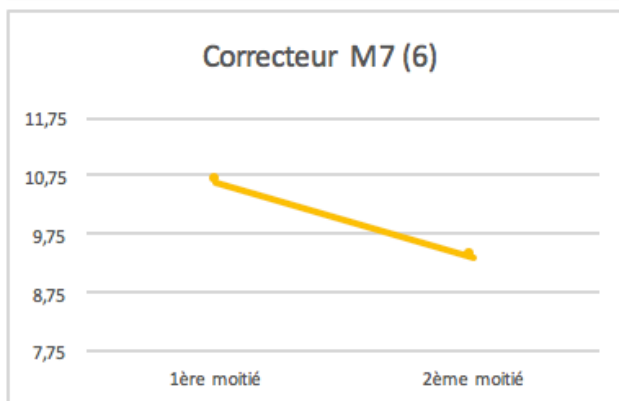
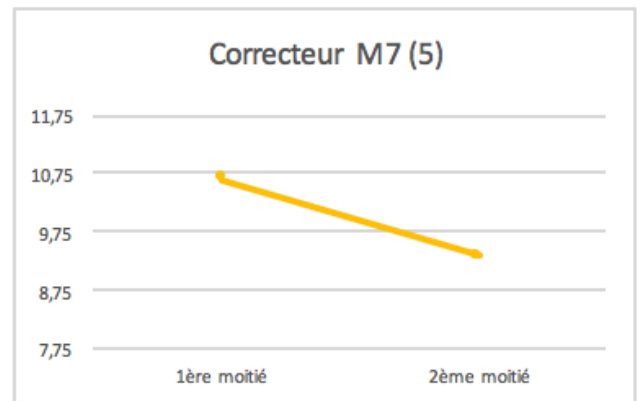
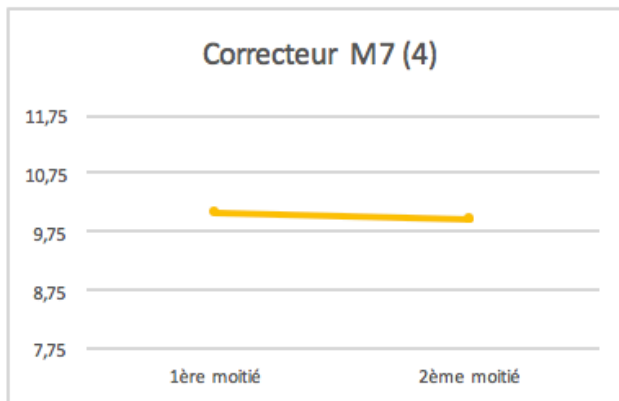
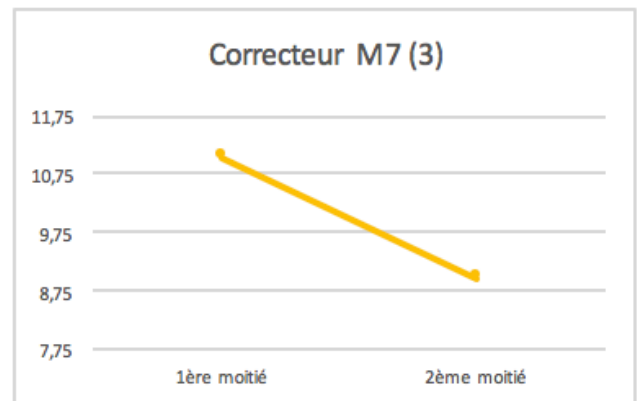
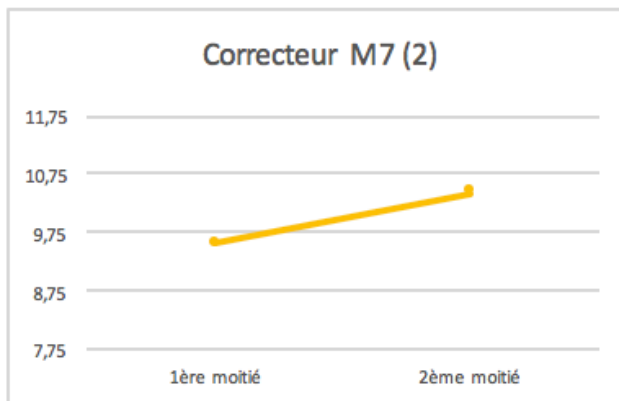
**Mode de correction : Exercice par exercice**

**Type de disciplines : Disciplines dites « normées »**

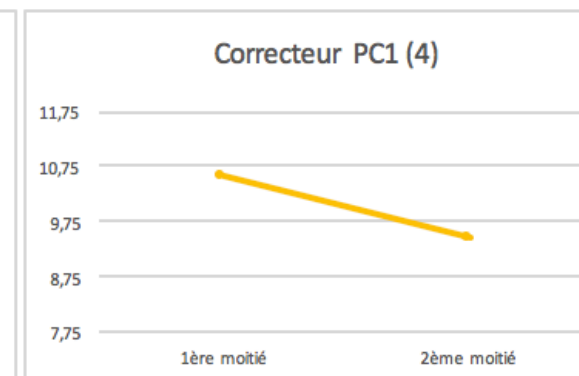
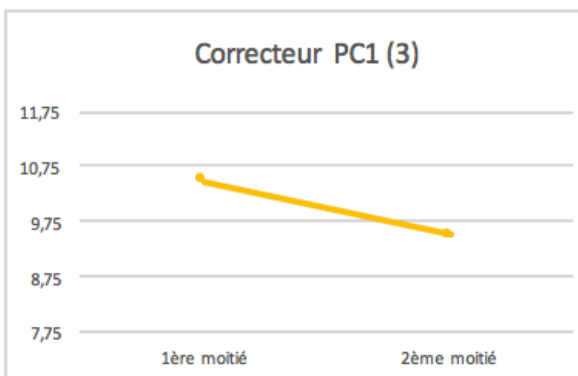
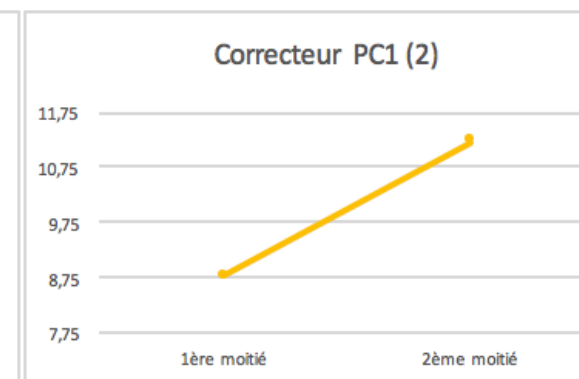
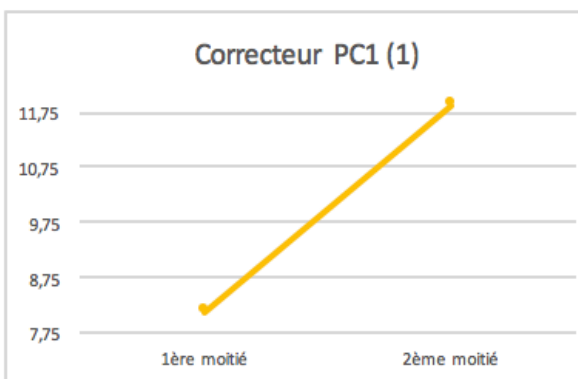
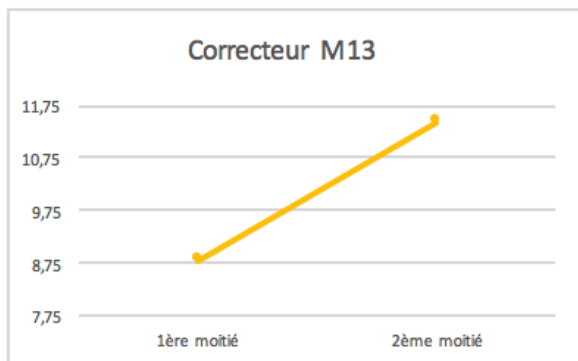
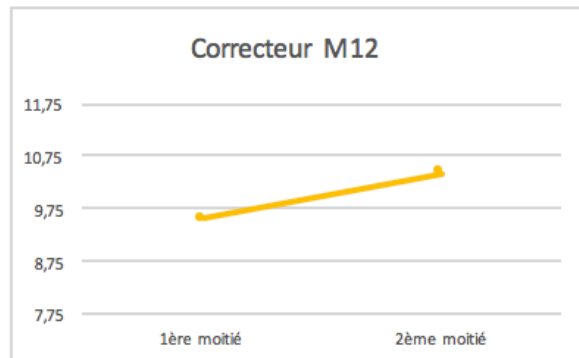
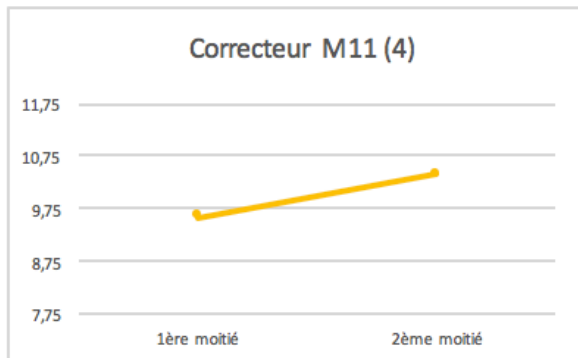
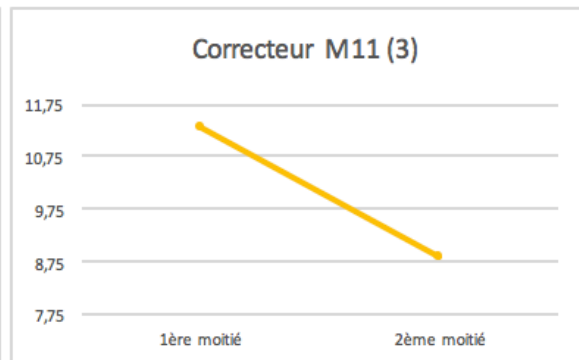
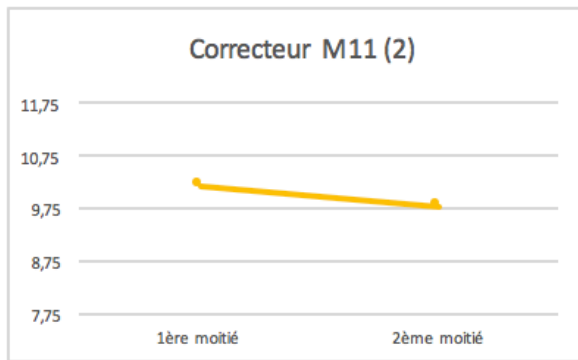




Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».



Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

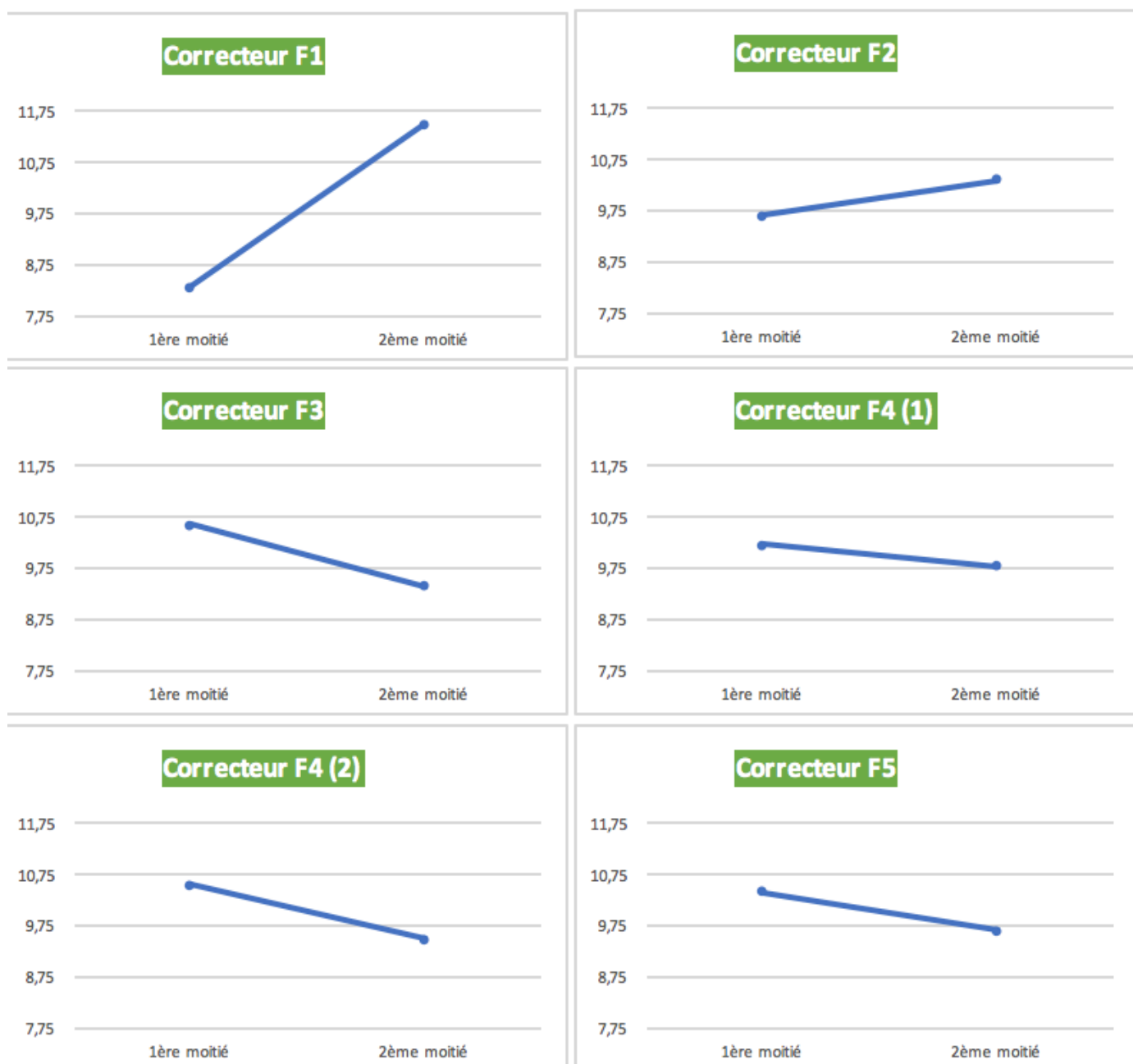


Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

**Méthode 1 : Séparation des séries en moitiés.**

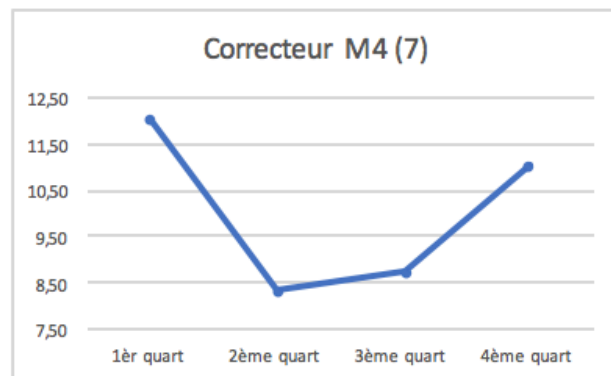
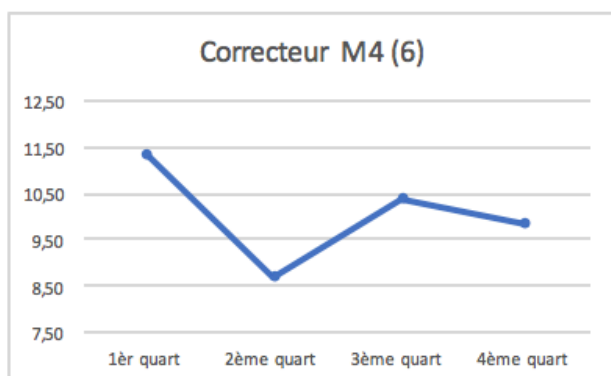
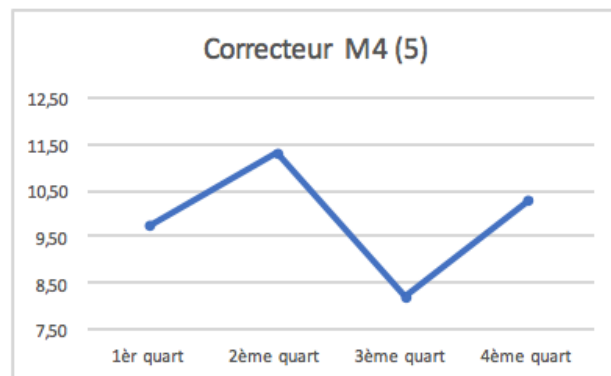
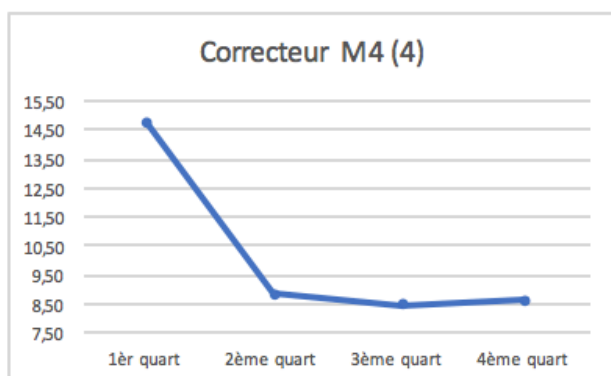
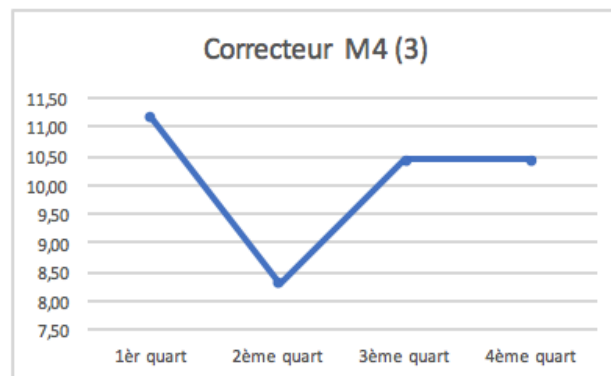
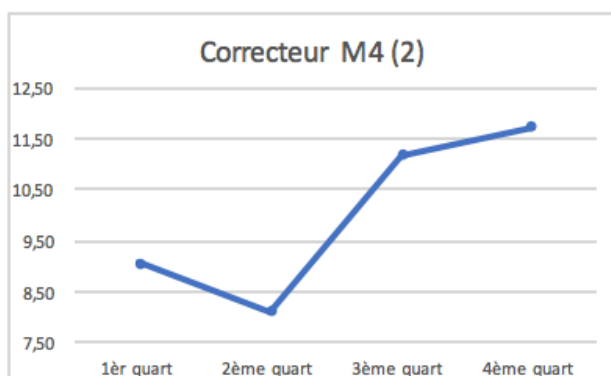
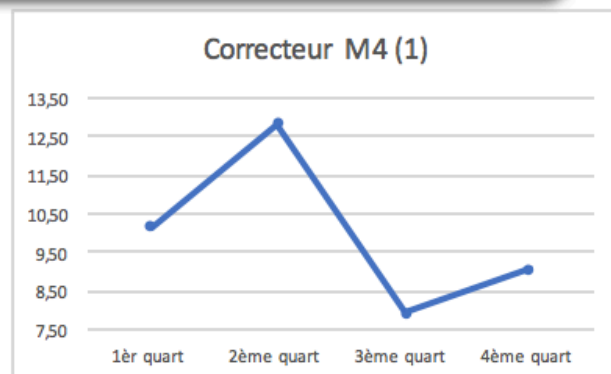
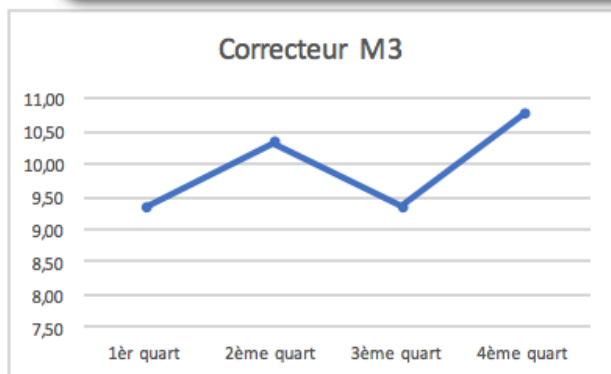
**Mode de correction : Copie par copie**

**Type de disciplines : Disciplines dites « moins normées »**

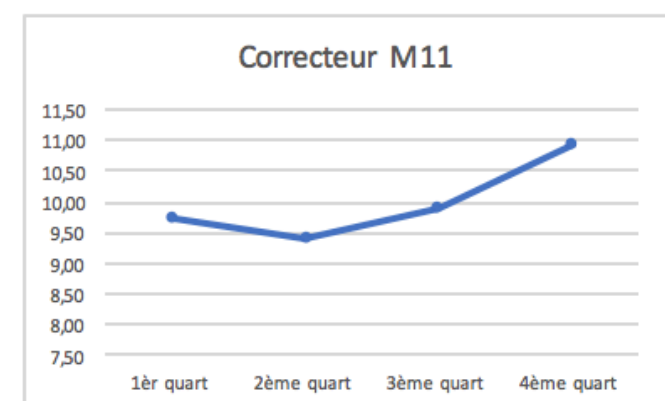
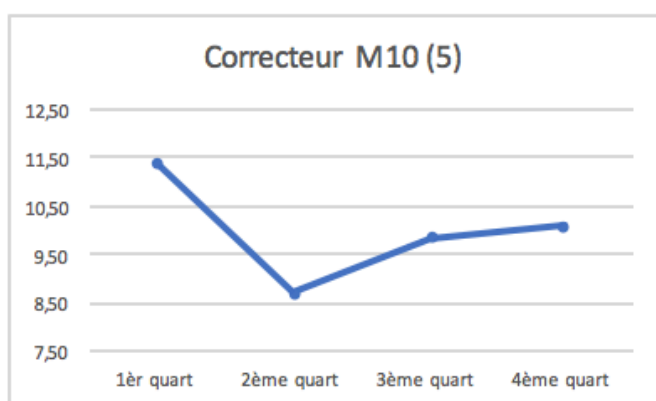
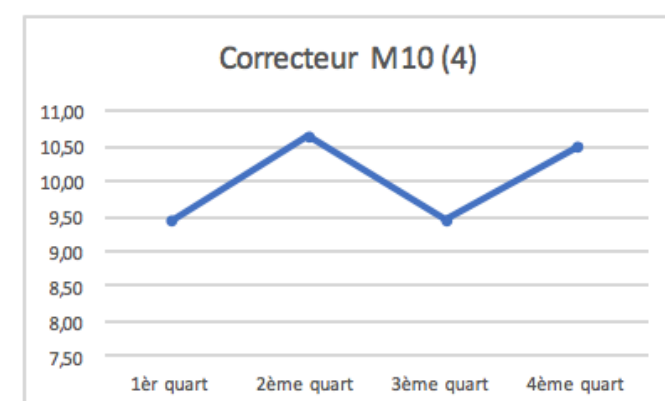
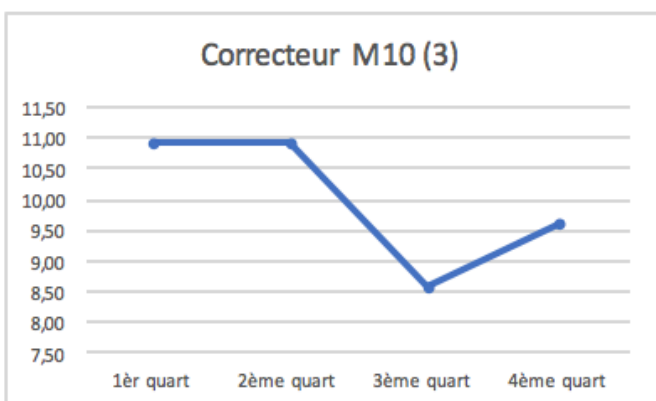
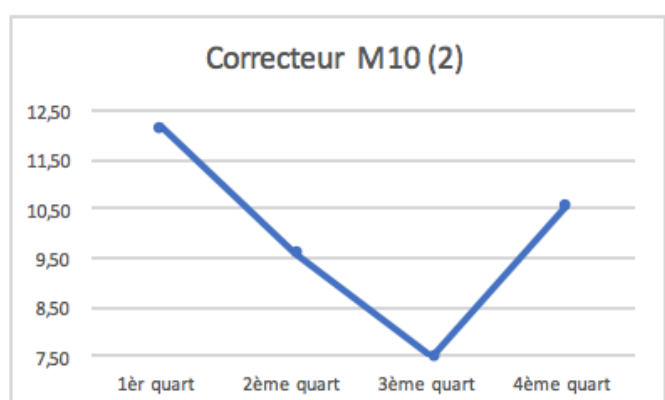
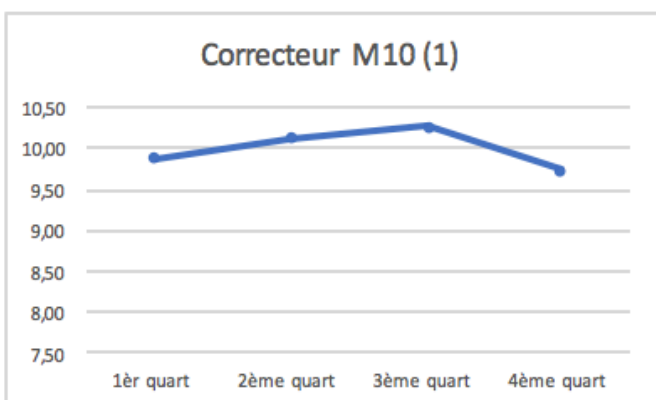
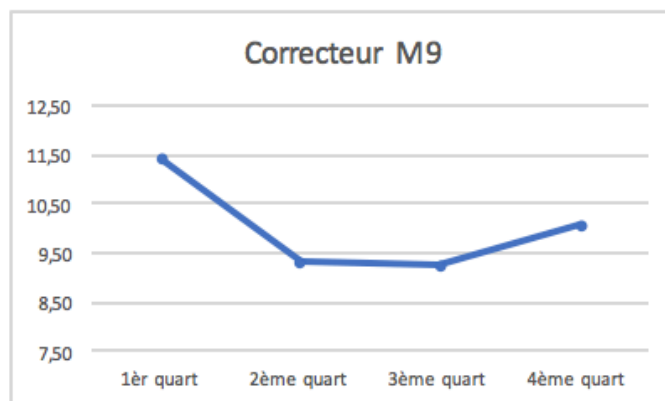
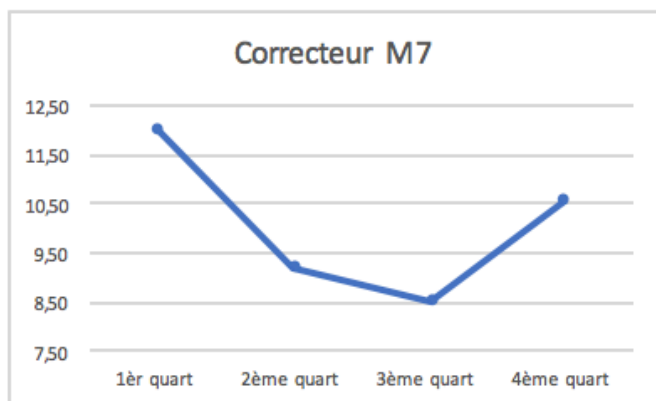


**Annexe 4 : Modélisations graphiques des données recueillies par séries selon la méthode 1 : traitement des données par sous-ensembles, séparation des séries en quarts.**

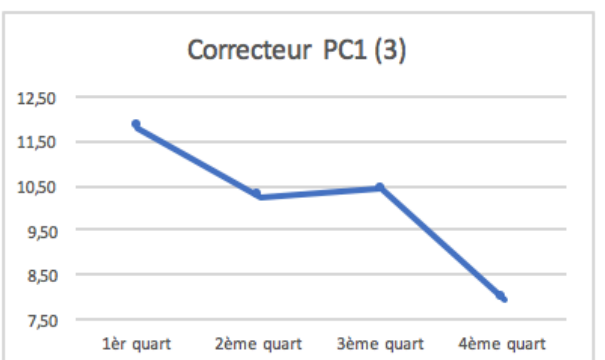
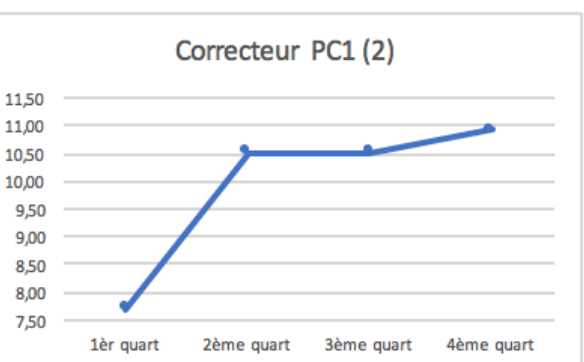
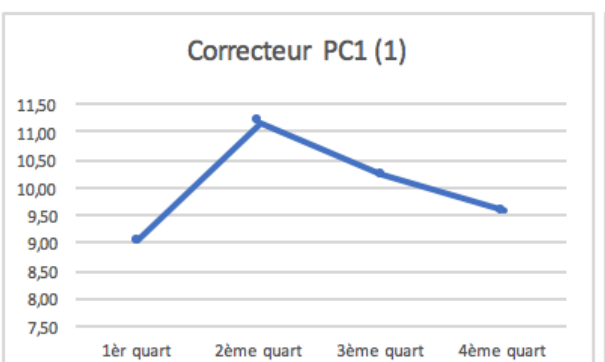
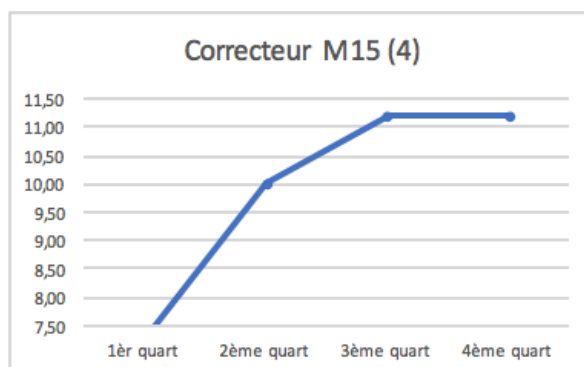
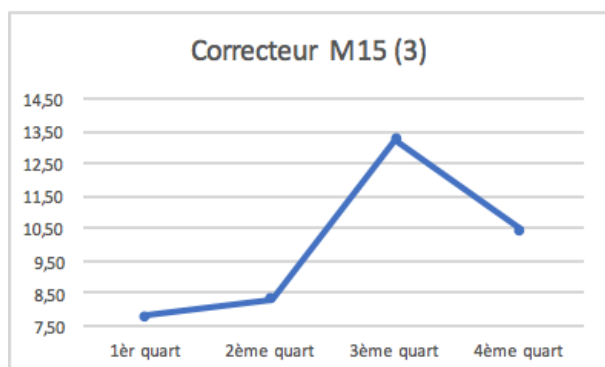
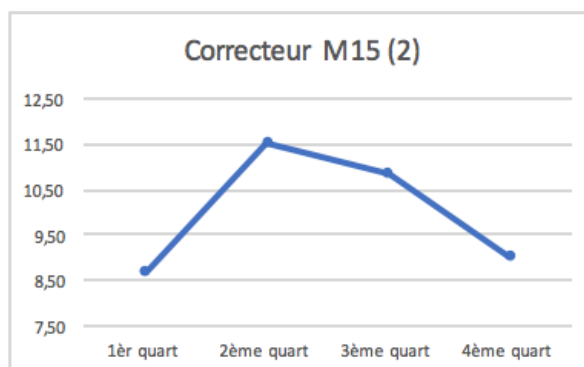
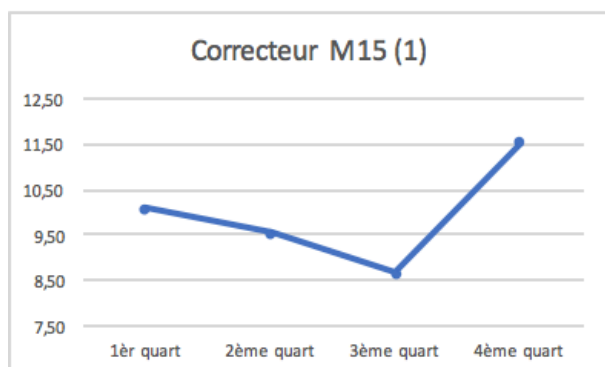
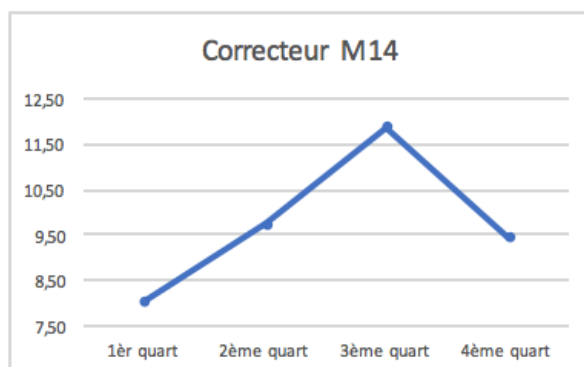
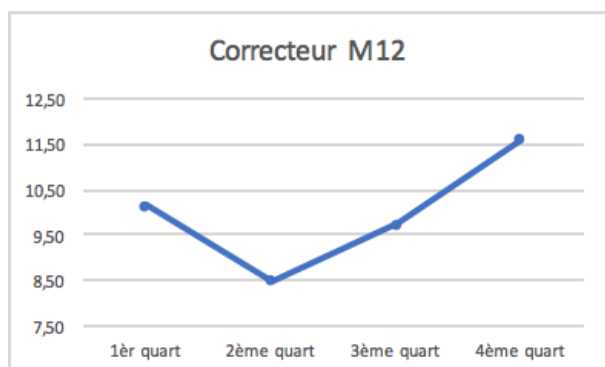
Méthode 1 : Séparation des séries en quarts.  
Mode de correction : Copie par copie  
Type de disciplines : Disciplines dites « normées »



Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».



Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

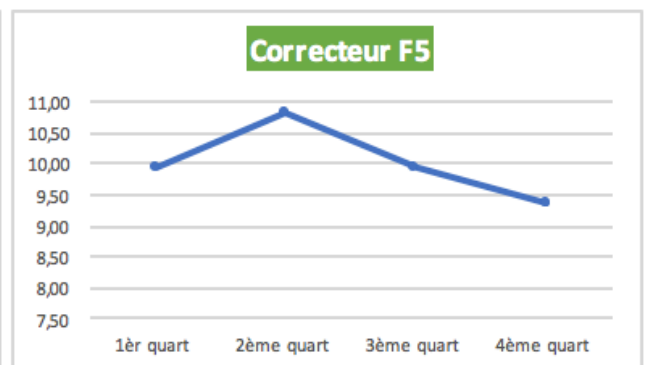
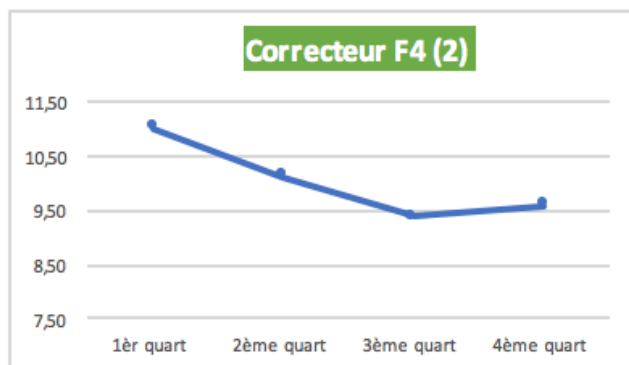
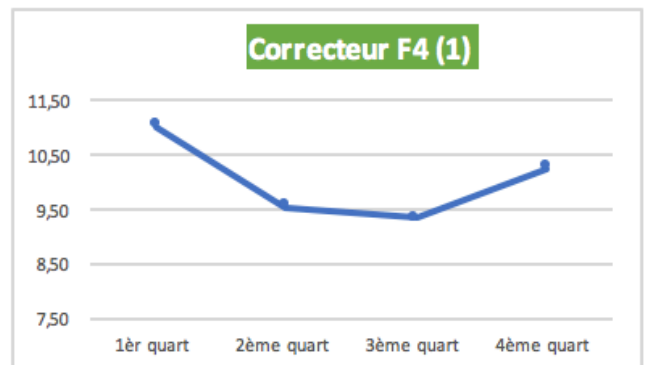
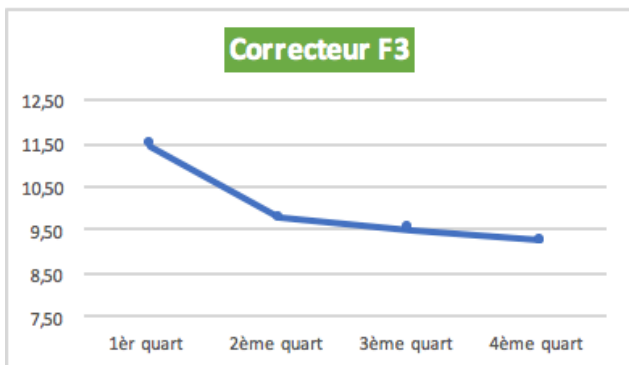
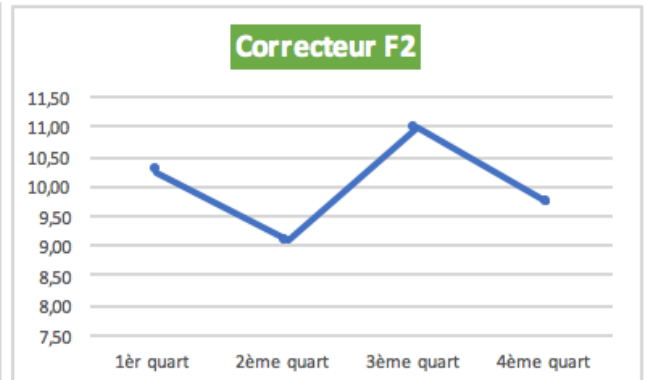
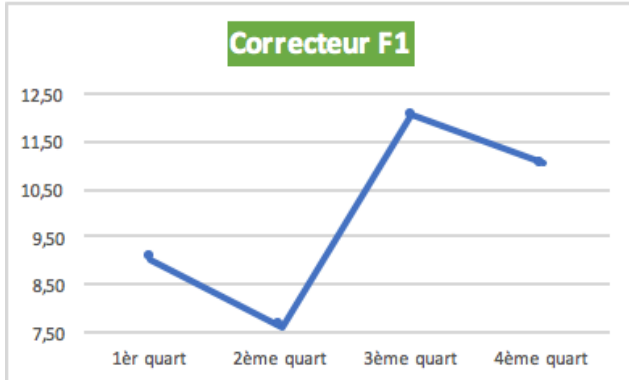


Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

**Méthode 1 : Séparation des séries en quarts.**

**Mode de correction : Copie par copie**

**Type de disciplines : Disciplines dites « moins normées »**

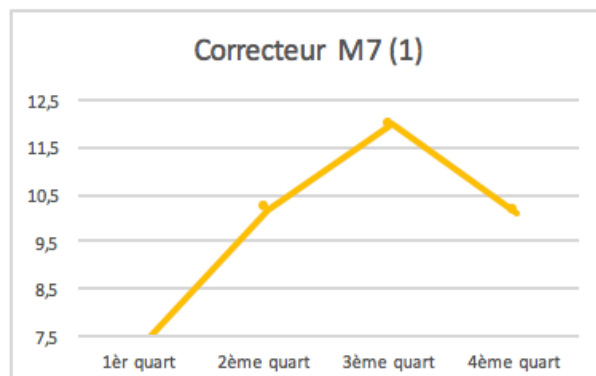
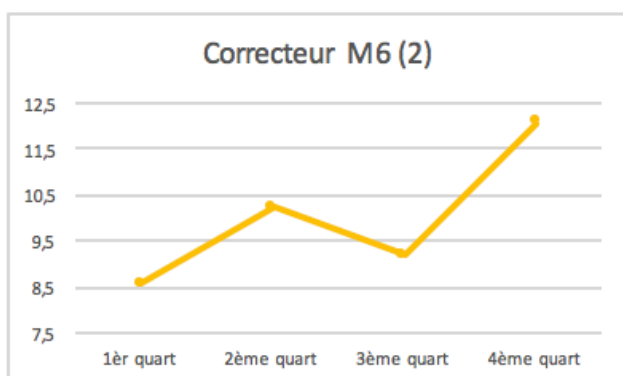
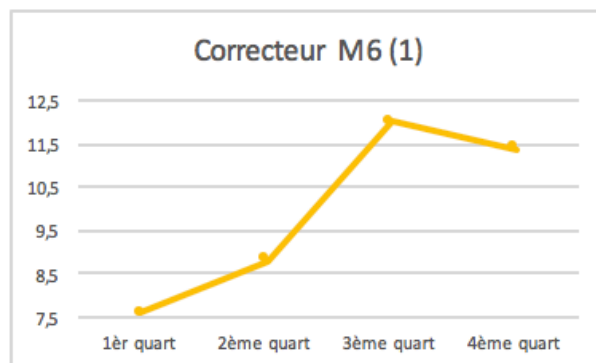
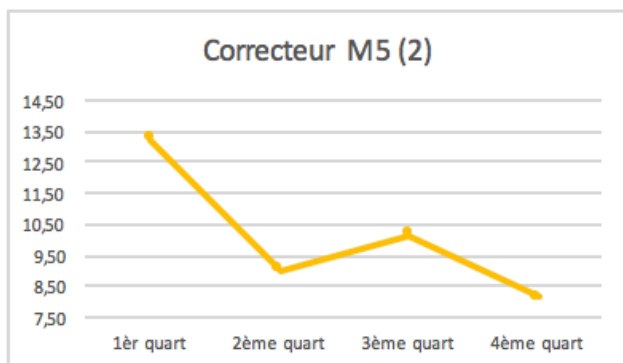
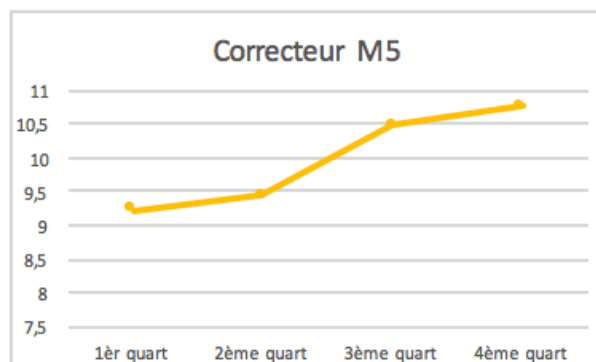
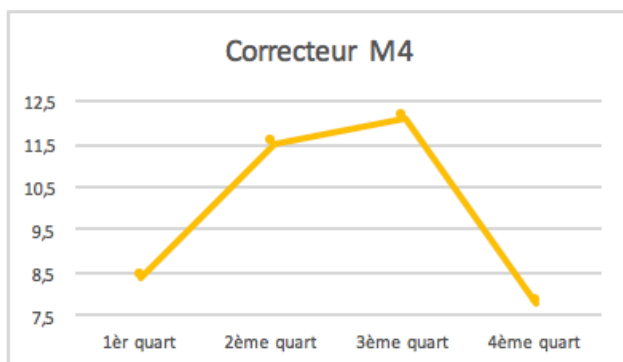
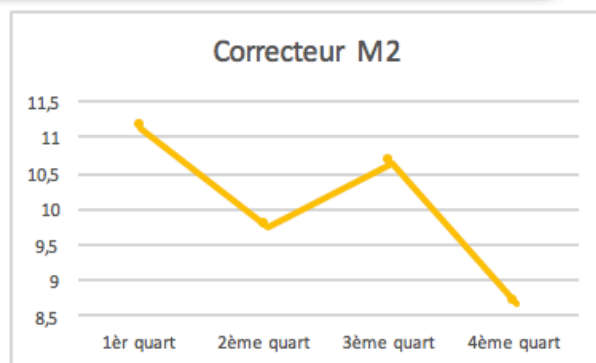
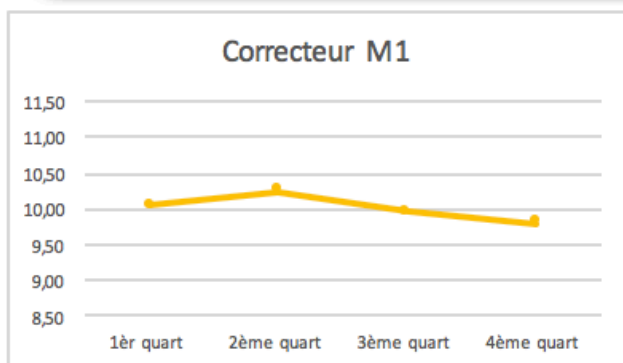


Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

**Méthode 1 : Séparation des séries en quarts.**

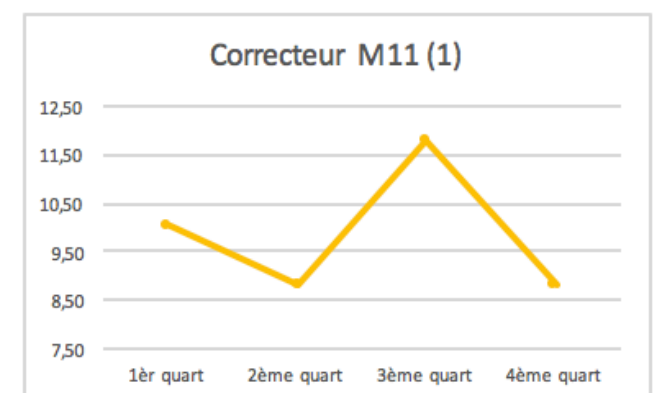
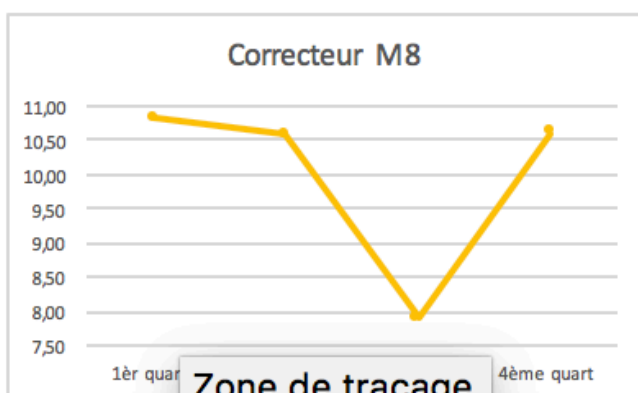
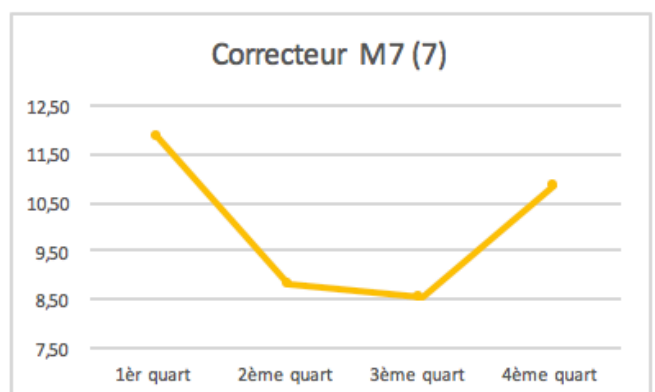
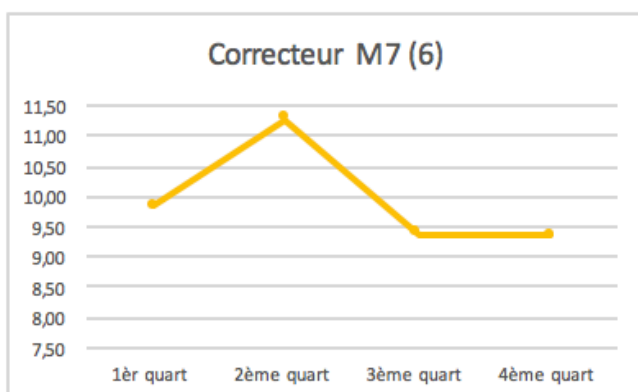
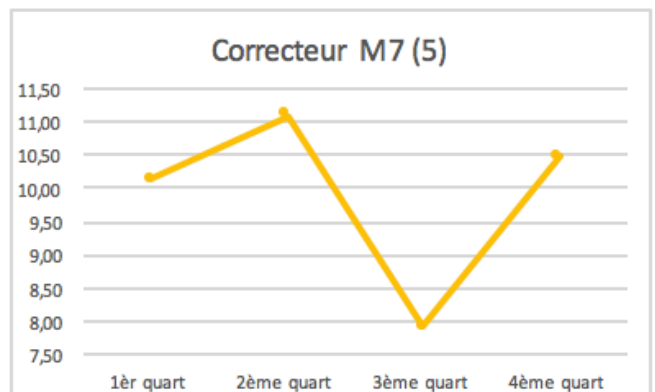
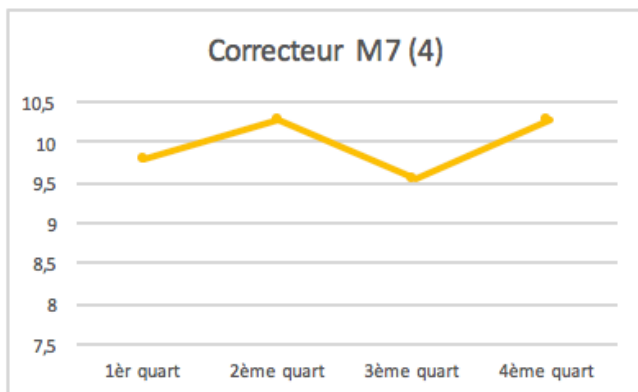
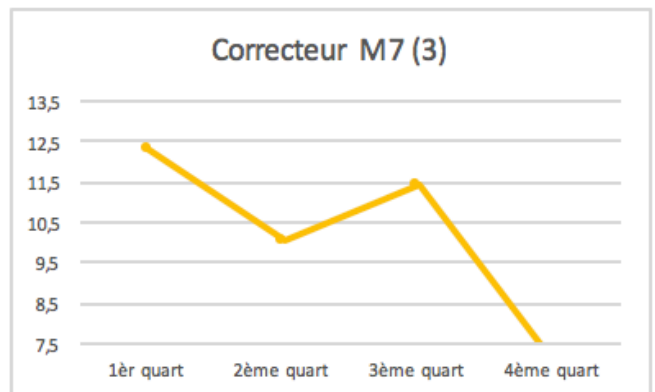
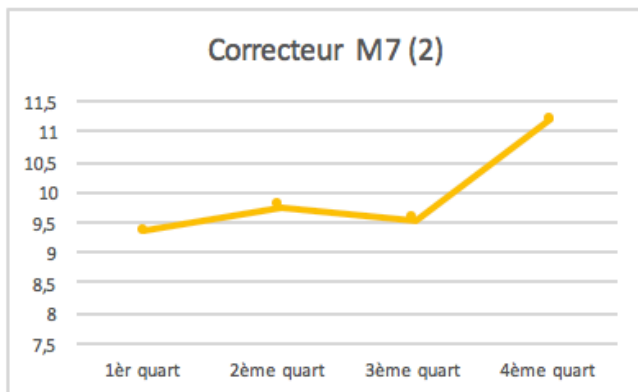
**Mode de correction : Exercice par exercice**

**Type de disciplines : Disciplines dites « normées »**



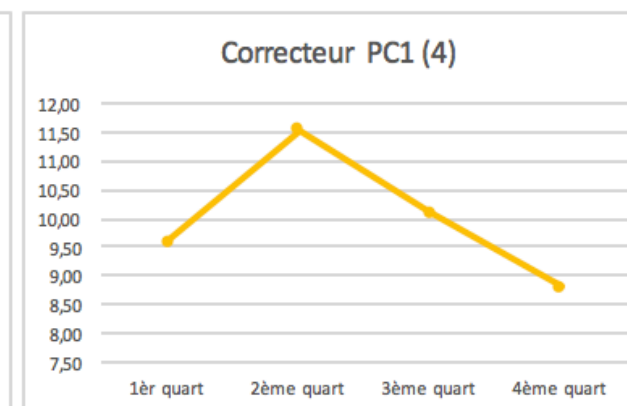
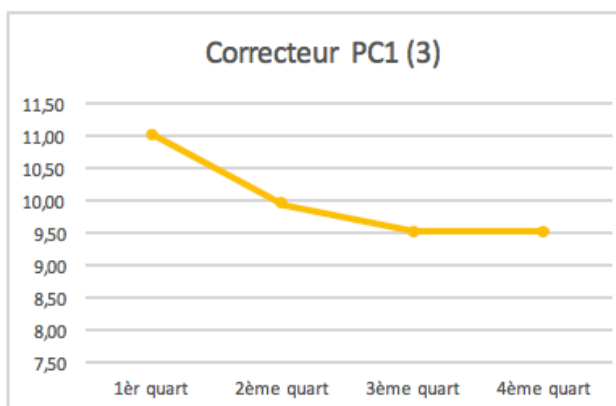
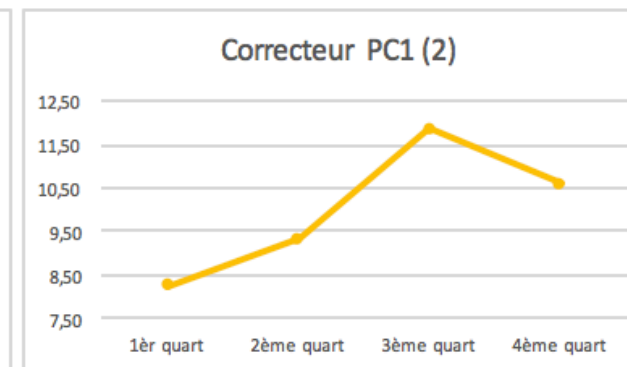
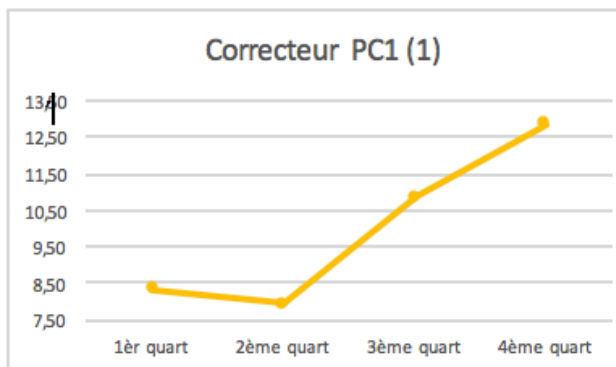
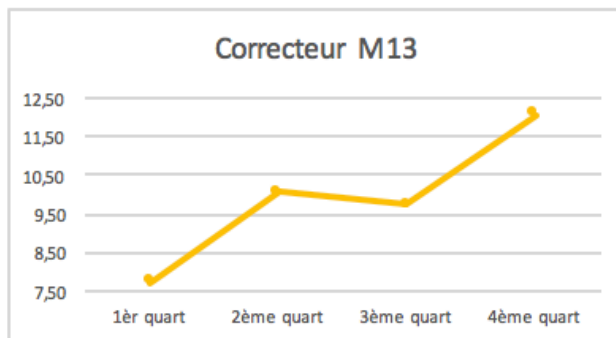
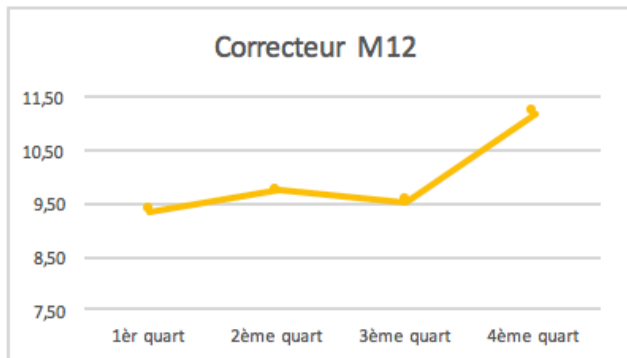
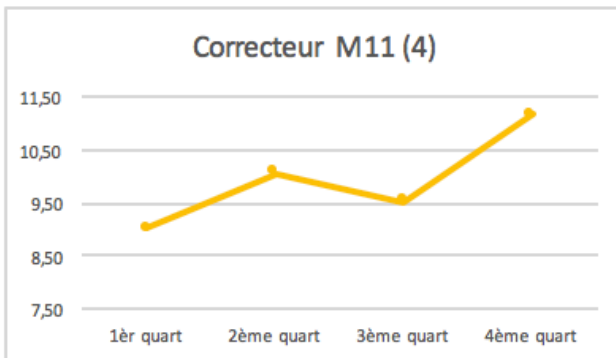
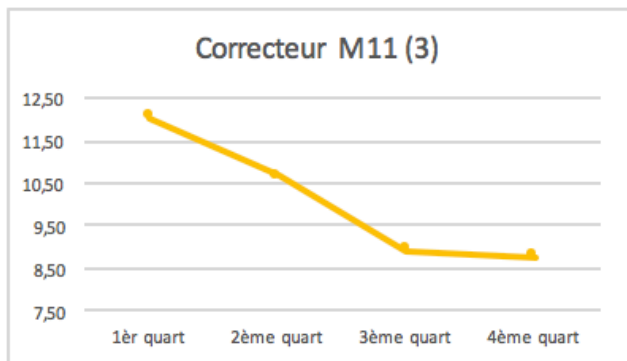
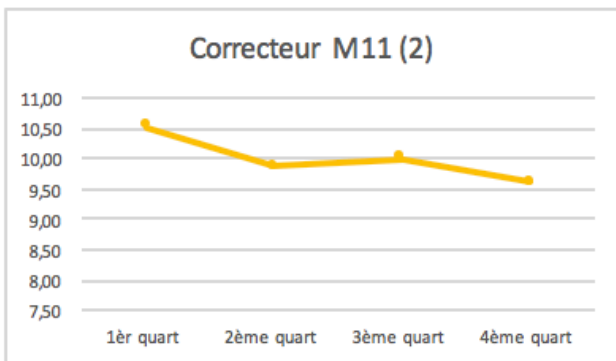


Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».



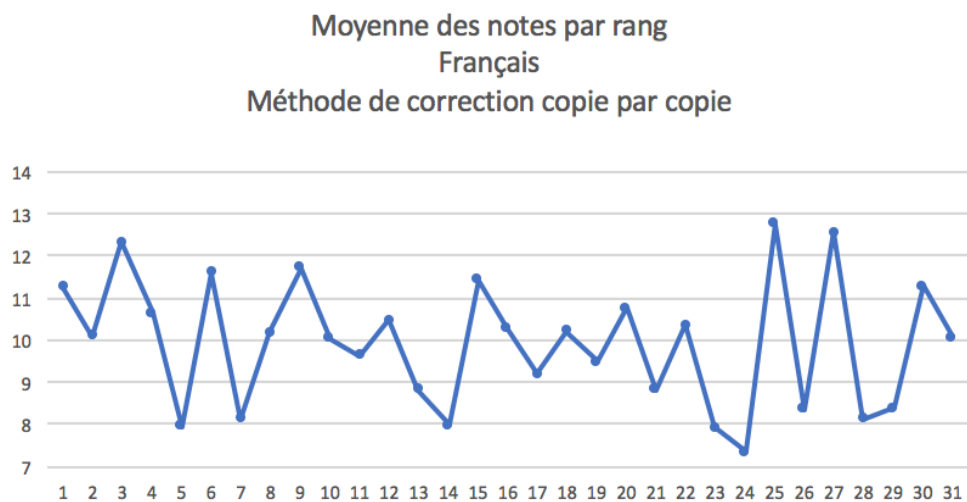
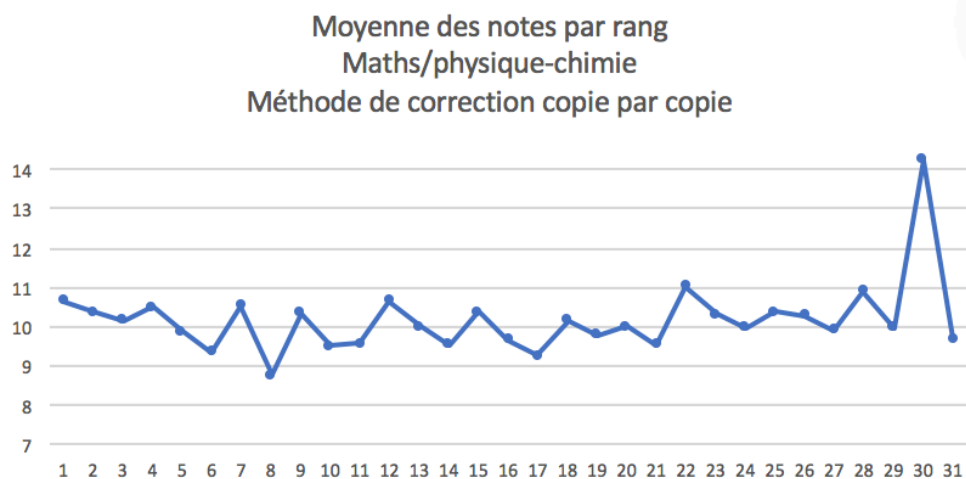
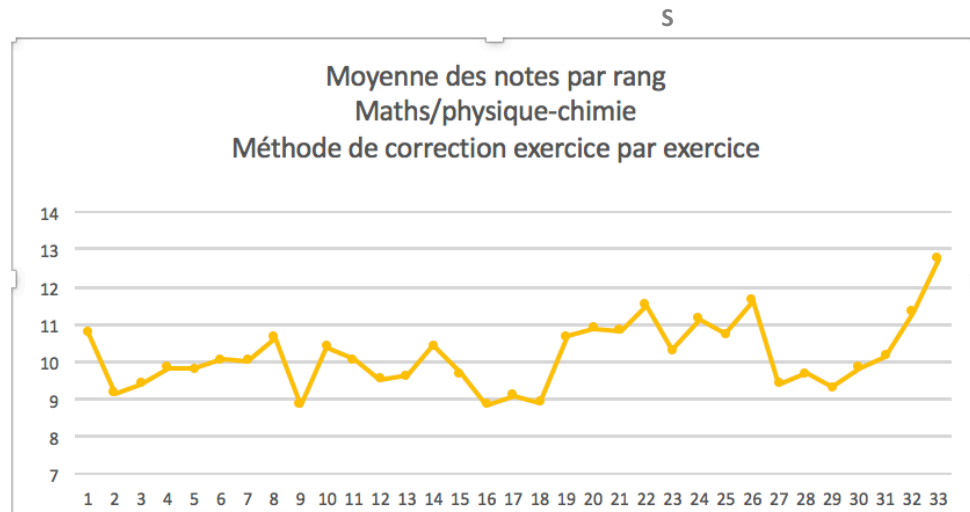
Zone de traçage

Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».



**Annexe 5 : Modélisations graphiques des données recueillies par types de disciplines et mode de correction suivi selon la méthode 2 : *traitement des données par rangs*.**

*Graphiques incluant l'ensemble des séries de notes.*



# Comparaison de l'influence de l'effet de position dans la distribution des notes entre les disciplines dites « normées » et « moins normées ».

Auteur : Amélie, COMBELLES

Directeur de mémoire : Yvan ABERNOT

Année : 2017

Nombre de pages : 37 pages

## Résumé :

Le travail de recherche présenté s'inscrit dans le domaine de la docimologie. En effet, cette étude vise à observer et à analyser l'influence de l'effet d'ordre (ou de position) lors de la correction de séries de copies, et ce, dans deux types de disciplines : les disciplines dites « normées » (types mathématiques, physique-chimie) et les disciplines dites « moins normées » (types français, philosophie).

56 séries de copies, issues des deux types de disciplines, ont été analysées et comparées entre elles afin de répondre à la question de recherche suivante :

*L'effet de position est-il moindre dans des disciplines dites « normées » (type mathématiques ou physique-chimie) par rapport à celui constaté dans d'autres disciplines dites « moins normées » (type français, philosophie) ?*

Malheureusement, à l'issue de l'étude, nous ne pouvons répondre à cette question en validant une de nos trois hypothèses de départ avec certitude. En effet, les données recueillies sont issues de conditions trop différentes pour pouvoir en tirer une analyse pertinente : ce sont des données de « terrain ». C'est-à-dire des notes réellement attribuées par des enseignants à leurs classes lors de la correction d'évaluation. De ce fait, en plus d'obtenir des séries de notes très hétérogènes en termes d'effectif, les notes en elles-mêmes ont pu être contaminées par d'autres effets parasites que celui étudié.

## Mots-clés :

Evaluation, docimologie, effet de position, disciplines dites « normées », disciplines dites « moins normées ».

## Abstract :

The enquiry introduced here comes within the domain of docimology. This study aims to observe and analyse the influence of ranking effects during correction of papers series, and, in two types of disciplines : the disciplines with norms (mathematics...) and disciplines with less norms (french, philosophy).

56 papers series, from two types of disciplines have been analysed and compared between them to answer to the next enquiry question : Position effect is it lesser in disciplines with norms compared to the others disciplines ?

Unfortunately, at the end of study, we can't answer at the question validating one of three hypothesis of beginning with certainty. These data collected are too different to can give an analyse relevant : these is data from "ground".

That is notes really give from teachers to classes during correction. Therefore, in addition to obtain series of very heterogeneous notes in terms of staff, this notes themselves were able to be contaminated by other effects parasites that studied.

Keywords : testing, docimology, ranking effects, disciplines "with norms", disciplines "with less norms".