



Master 2 « Métiers de l'Enseignement, de l'Education et de la Formation » pour l'enseignement agricole

Mémoire

Influence du rang de correction dans l'évaluation en mathématiques sur une échelle de 21 degrés (0-20).

Une étude simultanée de la position et de la distribution des notes de 94 séries de copies.

CANCET Emmanuelle

Ecole Nationale de Formation Agronomique



Jury :

Yvan Abernot, ENFA : Directeur de mémoire

Nina Asloum, ENFA : Examineur

Bernard Donnadiou, ENFA : Examineur

Mars 2016

Ecole Nationale de Formation Agronomique

BP 22687, 2 Route de Narbonne
31326 Castanet Tolosan cedex



Unité Mixte de Recherche « Education,
Formation, Travail, Savoirs »

Université de Toulouse II - Maison de la recherche
5, allées Antonio-Machado
F – 31 058 Toulouse Cedex 9



MEMOIRE EN SCIENCES DE L'EDUCATION
ANNEE DE TITULARISATION AU TITRE DE PCEA MATHEMATIQUES

**Influence du rang de correction dans
l'évaluation en mathématiques sur
une échelle de 21 degrés (0-20).**

***Une étude simultanée de la position et de la
distribution des notes de 94 séries de copies.***

CANCET Emmanuelle

Professeur stagiaire externe
PCEA Mathématiques

LEGTA Toulouse-Auzeville



Cité des Sciences Vertes
Établissement Public
APPRENTISSAGE • LYCÉE • ADULTES • EXPLOITATION
Toulouse-Auzeville

Directeur de mémoire: Y. Abernot

ENFA & UMR EFTS

Professeur

Remerciements

Je tiens à remercier Monsieur Abernot pour avoir accepté d'encadrer ce projet de mémoire, pour le suivi qu'il a réalisé et les conseils pertinents et nombreux, fournis tout au long de sa réalisation, ainsi que pour le prêt de nombreuses références bibliographiques.

Mes remerciements vont également vers Madame Chaput, formatrice en mathématiques à l'ENFA, qui m'a suggéré diverses pistes pour le traitement des données.

Enfin, un immense merci à tous les enseignants de mathématiques de l'enseignement agricole ayant accepté de collaborer au recueil des données.

En mathématiques, le plus trivial concret – ou voulu comme tel – comme la plus solide abstraction ont le même statut : celui d'une irréalité ou d'une sur-réalité qui sont celles d'un pays de rêve – ou de cauchemar, comme on voudra -, où les moutons peuvent avoir 32 pattes, les femmes mesurer 4,50 m, où 2 peut être égal à 6, où il faut empêcher 1 d'être égal à zéro, et dont on se demande comment les portes de 9 m tiennent encore sur leurs gonds, quelles sont ces mystérieuses allées sans retour au détour desquelles on rencontre petit a et petit b dont les liens de parenté passionneraient les ethnologues : du fait que petit a est la mère de petit b, mais que b ne peut être la mère de a, la relation est réflexive, ce qui fait que petit a et petit b peuvent être leur propre mère.

Baruk, S. (1985). *L'âge du capitaine: de l'erreur en mathématiques*. Paris: Éditions du Seuil. pp 18-19

Notice analytique

<p>Emmanuelle CANCEZ Professeur stagiaire externe PCEA Mathématiques <i>Ecole Nationale de Formation Agronomique (ENFA)</i></p>
<p>Type de travail et durée : Recherche en sciences de l'éducation – 5 mois –</p>
<p>Titre : Influence du rang de correction dans l'évaluation en mathématiques sur une échelle de 21 degrés (0-20).</p>
<p>Sous-titre : Une étude simultanée de la position et de la distribution des notes de 94 séries de copies.</p>
<p>Résumé : Cette recherche en docimologie s'intéresse à l'influence du rang de correction d'une copie lors de la correction d'une série de copie sur une échelle de 21 degrés, et ce, plus particulièrement dans l'évaluation des mathématiques. L'idée novatrice développée dans ce projet de recherche est d'étudier simultanément l'évolution de la moyenne et l'évolution de l'écart type (des notes) en fonction du rang de correction des copies. La principale hypothèse de travail autour de laquelle s'articule la problématique est l'hypothèse selon laquelle « au fur et à mesure de la correction la moyenne diminue et l'écart type augmente ». Au total, 94 séries de notes issues d'évaluations différentes ont fait l'objet d'une analyse afin de répondre aux différentes questions de recherche. Les effets d'ordres tels qu'ils ont été décrits dans le passé, n'ont pas été mis en évidence de façon claire. Bien que la principale hypothèse de travail n'ait pu être confirmée, elle ne peut être infirmée. En effet, les données expérimentales recueillies et analysées sont issues de corrections « réelles » ayant de ce fait pu être impactées à la fois par des effets parasites tels que les effets d'ordre, mais également par d'autres effets influant sur la validité d'une évaluation, comme par exemple, ceux de contamination, de stéréotypie et de halo.</p>
<p>Mots-clés : docimologie, validité de l'évaluation, effets d'ordre, mathématiques</p>
<p>Caractéristiques : 1 volume de 30 pages et 5 annexes, 16 références bibliographiques</p>
<p>Date de publication : 12 mars 2016</p>
<p>Directeur de mémoire : Y. Abernot, Professeur, ENFA et UMR-EFTS Toulouse</p>

Abstract

This docimological study deals with the influence of marking rank during the process of marking a set of copies and, more precisely, in the frame of mathematics evaluation on a 21 degree scale. The innovative idea developed in the frame of this research project is studying simultaneously the evolution of the following two parameters: average and standard deviation, with marking rank. The main working hypothesis around which problematic has been set up is "average values decrease and standard deviation values increase while the marking process goes on". A total of 94 series of marks coming from different evaluations have been analyzed in order to answer the questions asked in accordance with the problematic. Marking rank effects, as described in the past by previous studies, have not been highlighted in the current study. Although the main working hypothesis has not been corroborated, it cannot be rejected. Indeed, experimental data which have been collected and analyzed in the frame of the current study, comes from "actual" marking situations and as a consequence may have suffered others unwanted and/or side effects as marking rank effect but also other effects affecting measure validity as, for example, contamination effect, stereotype effect and halo effect.

Keys words : docimology, measure validity, ranking effects, mathematics

Sommaire

Remerciements	2
Notice analytique.....	3
Abstract.....	4
Table des figures.....	6
Table des tableaux.....	6
Table des abréviations	7
Table des annexes	7
Introduction	8
I. Cadre théorique et problématique	9
1. Contexte général : évaluation et docimologie.....	9
2. Le système évaluatif français : pratiques actuelles	9
3. Bref rappel de notions statistiques	10
4. Etude bibliographique	11
5. Questions de recherche.....	16
II. Méthodologie de recueil et traitement des données.....	17
1. Protocole expérimental : type de données recueillies.....	17
2. Protocole expérimental : recueil des données.....	17
3. Protocole expérimental : traitement des données	18
4. Protocole expérimental : analyse des données.....	20
III. Analyse réflexive des résultats de la recherche	21
1. Méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en quarts.....	21
2. Méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en moitiés.	24
3. Méthode 2 : moyennes et écarts types glissants.....	27
4. Méthode 3 : moyennes et écarts types par rang.	28
Conclusion et perspectives d'exploitation professionnelle	29
Bibliographie	30
Annexes	31

Table des figures

Figure 1: Résultats de correction de 3 copies de mathématiques par 150 correcteurs.	13
Figure 2: Ecart-types observés sur des séries de notes dans différentes matières.....	13
Figure 3: Mise en évidence par R. Weiss de l'effet de Halo, à l'occasion de la correction d'une production d'arithmétique.....	14
Figure 4: Les effets d'ordre (sens strict) dans le processus de correction d'une série de copies.....	15
Figure 5: Effets de contraste dans la correction d'une série de copies, en mathématiques...	15
Figure 6 : Profils « types » définis dans le cadre de l'analyse des données traitées par la méthode 1 dans le cas où les séries ont été séparées en quarts.....	21
Figure 7 : Profils « types » définis dans le cadre de l'analyse des données traitées par la méthode 1 dans le cas où les séries ont été séparées en moitiés.	24
Figure 8 : Exemple d'une série de données traitée par la méthode 2 avec $k=3$ (à gauche) et $k=6$ (à droite).....	27
Figure 9: Graphiques obtenus par la méthode 3 de traitement des données.	28

Table des tableaux

Tableau 1: Critères de caractérisation d'une évaluation : fidélité, sensibilité et validité.	11
Tableau 2 : Moyenne des notes attribuées à des versions anglaises, en fonction du sous-paquet considéré, selon l'ordre d'évaluation. Données d'étude de Bonniol, J.J. (1965).	17
Tableau 3 : Provenance (correcteur) des séries de copies prises en compte dans la phase de traitement des données en fonction du type de correction leur correspondant.....	18
Tableau 4 : Méthodes de traitement des données.	19
Tableau 5 : Résultats d'analyses des graphiques correspondant à une correction de type "exercice par exercice" (méthode 1 - séparation des séries en quarts).....	22
Tableau 6 : Résultats d'analyses des graphiques correspondant à une correction de type "copie par copie" (méthode 1 - séparation des séries en quarts).	22
Tableau 7 : Confrontation des résultats obtenus en fonction du type de correction (méthode 1 - séparation des séries en quarts).	23
Tableau 8 : Résultats d'analyses des graphiques correspondant à une correction de type "exercice par exercice" (méthode 1 - séparation des séries en moitiés).	25
Tableau 9 : Résultats d'analyses des graphiques correspondant à une correction de type "copie par copie" (méthode 1 - séparation des séries en moitiés).....	26
Tableau 10 : Confrontation des résultats obtenus en fonction du type de correction (méthode 1 - séparation des séries en moitiés).	26

Table des abréviations

CRDP	Centre Régional de Documentation Pédagogique
EA	Enseignement Agricole
EFTS	Education, Formation, Travail, Savoirs
ENFA	Ecole Nationale de Formation Agronomique
LEGTA	Lycée d'Enseignement Général et Technologique Agricole
PCEA	Professeur Certifié de l'Enseignement Agricole
UMR	Unité Mixte de Recherche

Table des annexes

Annexe 1: Formulaire de recueil de données	32
Annexe 2: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en quarts.	35
Annexe 3: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en moitiés.	45
Annexe 4: Planches de résultats obtenus par la méthode 2 : moyennes et écarts types glissants.	55
Annexe 5: Résultats obtenus par la méthode 3 : moyennes et écarts types par rang.	65

Introduction

Les travaux effectués dans le cadre du projet de recherche en Sciences de l'Education ici reporté, ont pour but de caractériser davantage un phénomène de docimologie influant sur la validité d'une évaluation, à savoir, l'effet d'ordre de correction. En d'autres termes, cette recherche s'intéresse à l'influence du rang de correction d'une copie lors de la correction d'une série de copie sur une échelle de 21 degrés (échelle ordinale de 0 à 20), et ce, plus particulièrement dans l'évaluation des mathématiques. Il est important de préciser que seul le cas de l'évaluation interne, c'est-à-dire, l'évaluation pratiquée par la personne qui dispense l'enseignement, n'a été abordé dans le cadre de cette recherche. L'idée novatrice développée dans ce projet de recherche est d'étudier simultanément l'évolution de la moyenne et l'évolution de l'écart type (des notes) en fonction du rang de correction des copies.

La recherche s'est chronologiquement articulée en trois étapes qui sont décrites dans les trois parties de ce rapport, respectivement. L'étude bibliographique du cadre théorique ainsi que des connaissances actuelles sur le sujet ayant permis de dégager la problématique de recherche, font l'objet de la première partie. La partie suivante s'attache à décrire la méthodologie de traitement et de recueil des données, depuis le choix du type de données recueillies, jusqu'aux méthodes de traitement et d'analyse de celles-ci. Au total, 96 séries de notes issues d'évaluations différentes ont été reçues, parmi lesquelles 94 ont fait l'objet d'un traitement et d'une analyse afin de répondre aux différentes questions de recherche. Enfin, les résultats de l'analyse des données en tant que telle sont consignés et commentés dans la troisième partie de ce mémoire.

I. Cadre théorique et problématique

1. Contexte général : évaluation et docimologie

La docimologie, science des examens, est plus particulièrement définie par De Landsheere, G. (1980), comme étant l'étude systématique des examens et en particulier des systèmes de notation et du comportement des examinateurs et des examinés. C'est H. Pieron, en 1963, qui institua cette discipline orientée vers l'étude systématique des examens et qui lui donna le nom de docimologie, du grec *dokimê*, épreuve et *logos*, science.

On peut lire dans De Ketele (1982), que dans ses premiers développements, la docimologie a revêtu un caractère essentiellement négatif en critiquant les modes de notation et en se focalisant principalement sur l'étude du manque de fidélité et de validité des examens. La docimologie, au sens actuel et élargi du terme, est la discipline qui a pour objet l'étude des systèmes d'évaluation en éducation et dont la finalité réside dans l'amélioration des systèmes d'évaluation.

En docimologie, le processus fondamental est l'évaluation lui-même relié à d'autres processus fondamentaux comme l'observation, la notation et la mesure, entre autres (De Ketele. J-M., 1982). En effet, selon cet auteur, il ne peut y avoir d'évaluation sans observation ; noter signifie attribuer selon un barème fixé une appréciation synthétique objective ou subjective traduisant l'évaluation d'une performance ou d'une conduite, la notation étant une forme possible d'évaluation; mesurer signifie attribuer des nombres à des « choses » selon des règles bien déterminées en respectant le principe d'isomorphisme (similarité entre les propriétés des objets mesurés et les propriétés du système de mesure). La mesure n'est pas nécessairement un acte évaluatif, de même que l'évaluation ne suppose pas nécessairement le recours à une mesure.

Concernant plus précisément le terme évaluation, Abernot Y. (1996), parle de paradoxe du fait que l'évaluation serait à la fois une mesure et une appréciation. Selon De Ketele (1982), évaluer signifie *examiner le degré d'adéquation entre un ensemble d'informations et un ensemble de critères adéquats à l'objectif fixé, en vue de prendre une décision.*

2. Le système évaluatif français : pratiques actuelles

L'acte évaluatif est un devoir inscrit dans le Code de l'Education¹, auquel doivent répondre les enseignants, quelque soit le niveau au sein duquel ils exercent ; depuis le plus jeune âge (école maternelle), jusqu'aux études supérieures, les élèves doivent être régulièrement évalués :

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. (...) Article L311-1, Chapitre 1er : Dispositions communes.

Les programmes définissent, pour chaque cycle, les connaissances et les compétences qui doivent être acquises au cours du cycle et les méthodes qui doivent être assimilées. (...) Article L311-3, Chapitre 1er : Dispositions communes.

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

¹ <http://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006071191>

responsabilité du directeur ou du chef d'établissement. (...) Article L311-7, Chapitre Ier : Dispositions communes.

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. (...) Les enseignants tiennent informés les parents d'élèves et les aident à suivre la scolarité de leurs enfants. (...) Article L912-1, Chapitre II : Dispositions propres aux personnels enseignants.

Quand bien même l'acte évaluatif revêt un caractère obligatoire, aucun système ou échelle de notation en particulier ne sont imposés par le Code de l'Education. Lorsqu'on se réfère à l'évaluation en France, on est - assez naturellement - amené à penser à la classique échelle ordinale sur 21 degrés (échelle de notation de 0 à 20). Cependant, de nos jours, d'autres systèmes ou échelles de notations sont communément utilisés: e.g. échelle de notation nominale du type « A, B, C, D, E » ; certains enseignants n'hésitent pas, d'ailleurs, à mélanger les systèmes de notation au sein de leurs pratiques.

3. Bref rappel de notions statistiques

On parle de série statistique, lorsque, pour une population et un caractère donnés, on connaît l'occurrence des valeurs, encore appelées modalités dudit caractère, et ce, pour chaque individu constituant la population d'étude.

Moyenne et écart type

La moyenne est un indicateur statistique dit « de position », alors que l'écart type est un indicateur statistique dit « de dispersion ». L'écart type correspond littéralement à la racine carrée de la variance qui elle, n'est autre que la moyenne des écarts quadratiques à la moyenne de la série. L'écart type est donc l'écart quadratique moyen des valeurs de la série statistique autour de la valeur moyenne ; il mesure la dispersion d'une série de valeurs autour de la valeur moyenne de la série statistique considérée. On utilise plus généralement l'écart type pour des raisons d'homogénéité d'unité. En effet, on vérifie facilement que l'unité de la variance est égale au carré de l'unité dans laquelle est exprimé le caractère étudié, alors que l'écart type est homogène à l'unité dans laquelle est exprimé le caractère étudié.

La moyenne, notée mu (μ), et l'écart type, noté sigma (σ), se calculent de la façon suivante :

$$\mu = \frac{\sum_{i=1}^p n_i x_i}{N} \quad \text{où :}$$

n_i : nombre d'occurrence/effectif de chaque note

$$\sigma = \sqrt{\frac{\sum_{i=1}^p n_i (x_i - m)^2}{N}} \quad \begin{array}{l} x_i : \text{les valeurs ou modalités de la série statistique} \\ N : \text{effectif total de la série statistique } (N = \sum_{i=1}^p n_i) \end{array}$$

Dans le contexte de la recherche ici présentée, le paramètre N correspond au nombre de copies considérées dans le calcul de l'écart type ; les valeurs x_i correspondent aux 21 valeurs de l'échelle de note (de 0 à 20).

Centrage et réduction des valeurs

Afin de comparer aisément des notes attribuées par différents correcteurs et de s'affranchir des phénomènes de sévérité-générosité et des écarts de niveau qu'il peut y avoir d'un établissement à l'autre ou d'une classe à l'autre, un centrage et une réduction des notes peuvent être réalisés dans le but de ne conserver que les écarts relatifs entre les notes

attribuées aux copies d'une même série. Dans les faits, le centrage et la réduction consistent à ramener la moyenne et l'écart type des notes de chaque série de copies, à une valeur donnée. Le développement théorique permettant d'opérer un centrage et une réduction des notes d'une série de copie est donné ci-dessous :

Soit m , la « nouvelle » moyenne de la série de copie et s le « nouvel » écart type de la série de copies.

Soit X , la variable aléatoire discrète prenant pour valeur la note initiale attribuée à une copie. L'univers image de X est : $X(\Omega) = \{0; 1; 2; 3; \dots; 19; 20\}$. L'espérance mathématique de X , notée $E(X)$ est égale à μ et la variance de X , notée $V(X)$ est égale à σ^2 .

Soit $Z = aX + b$, la variable aléatoire discrète prenant pour valeur la note centrée réduite de la copie. Les propriétés de l'espérance mathématique et de la variance nous permettent de résoudre le système suivant et d'en déduire les expressions des inconnues a et b :

$$\begin{aligned} \begin{cases} E(Z) = m \\ V(Z) = s^2 \end{cases} &\Leftrightarrow \begin{cases} E(aX + b) = m \\ V(aX + b) = s^2 \end{cases} \Leftrightarrow \begin{cases} aE(X) + b = m \\ a^2V(X) = s^2 \end{cases} \Leftrightarrow \begin{cases} a\mu + b = m \\ a^2\sigma^2 = s^2 \end{cases} \Leftrightarrow \begin{cases} b = m - a\mu \\ a^2 = \frac{s^2}{\sigma^2} \end{cases} \\ \Rightarrow \begin{cases} b = m - a\mu \\ a = \frac{s}{\sigma} \end{cases} &\Rightarrow \begin{cases} b = m - \frac{s}{\sigma}\mu \\ a = \frac{s}{\sigma} \end{cases} \Rightarrow \boxed{Z = \frac{s}{\sigma}X + \left(m - \frac{s\mu}{\sigma}\right)} \end{aligned}$$

4. Etude bibliographique

La docimologie classique s'intéresse aux problèmes de fidélité, de validité et de sensibilité de l'évaluation (Tableau 1). *Il s'agit de trouver les lois qui rendraient compte des problèmes de la « fidélité » des notes (aboutir au même résultat quelque soit le nombre de passations ou de correcteurs, de la « validité » des examens (n'évaluer que ce qui est affiché), et de la « sensibilité » des outils d'évaluation (utiliser harmonieusement les échelles de mesure).*²

On pourrait penser à tort que dans une discipline normée, c'est à dire constituée en systèmes de règles, comme le sont les mathématiques, les évaluations sont toujours fidèles, valides et sensibles.

FIDELITE	SENSIBILITE	VALIDITE
<p>Ce critère traite de la stabilité dans le temps de la note attribuée à une copie par un même correcteur (auto-fidélité ou encore variation intra-individuelle) ou de la stabilité de celle-ci par rapport à différents correcteurs (variation inter-individuelle).</p> <p>En toute théorie, une évaluation fidèle devrait amener au même résultat, quelque soit l'heure et le jour de la correction et quelque soit le correcteur.</p>	<p>Ce critère reflète la problématique d'adéquation entre l'échelle de note ou du système de mesure utilisé et la complexité, la variabilité des productions d'élèves devant être évaluées.</p> <p>Un outil d'évaluation sensible utilise une échelle de mesure ou de note en adéquation avec le degré de complexité et les variations des productions d'élèves devant être évaluées.</p>	<p>Ce critère traite de la correspondance entre ce qui est à évaluer et de ce qui est effectivement évalué. Lorsqu'on s'intéresse à la validité d'un outil d'évaluation, on cherche à répondre à la question : ce que l'on souhaite mesurer est-t-il correctement mesuré ?</p> <p>Un outil d'évaluation valide mesure ce que l'on prétend mesurer et rien d'autre.</p>

Tableau 1: Critères de caractérisation d'une évaluation : fidélité, sensibilité et validité.

² Citation extraite de Bonniol J.J. et Vial, M. (1997)

Néanmoins, les études sur le sujet mirent en exergue que même en mathématiques, discipline où l'on pourrait s'attendre à une très bonne reproductibilité des résultats de correction, l'évaluation présente des défauts de fidélité, sensibilité et validité.

Abernot Y. (1996), De Landsheere G., (1980), Bonboir A., (1972), Bonniol et Vial (1997), font un état des lieux et discutent, dans leurs ouvrages respectifs, de différentes causes influant la fidélité, la sensibilité et la validité d'une évaluation.

Ci-dessous, sont exposés quelques exemples de défauts de fidélité, de sensibilité et de validité, et ce, dans le cadre de l'évaluation en mathématiques.

Fidélité de l'évaluation en mathématiques

Dans le cadre de l'enquête Carnegie³, Laugier H et Weinberg D. (1936), travaillèrent sur les résultats du baccalauréat⁴. Ils s'intéressèrent aux variations inter individuelles : selon leurs résultats, en mathématiques, plus de deux points séparent la moyenne du correcteur le plus « sévère » de celle obtenue par le plus « généreux » : 7,01/20 contre 9,11/20). Pour une même copie, l'écart de notation maximal relevé est de 9 points.

Par ailleurs, les mêmes auteurs s'intéressèrent également à la notion de « valeur vraie » d'une note, qu'ils définirent plus ou moins en ces termes : la valeur vraie d'une note est égale à la moyenne des notes attribuées par un ensemble de correcteurs dont l'effectif assure que cette valeur moyenne soit stable. Pour une copie de mathématiques, 13 correcteurs seraient nécessaires afin d'assurer la stabilité de la moyenne des notes attribuées, et ce, pour un coefficient de fidélité égal à 0,99 (calcul basé sur la formule de Spearman-Brown⁵ : pour de plus amples détails sur cette étude et sur les développements théoriques employés, se reporter à Laugier H. et Weinberg D. (1936) et Langouet et Porlier (1994), respectivement).

Bien que datant un peu, voilà un résultat surprenant méritant d'être cité: R. Duquenne (1967) répartit des correcteurs en deux groupes : ceux s'affirmant « sévères » et ceux se considérant « indulgents ». Quatre devoirs identiques (matière non renseignée) furent alors corrigés par chaque correcteur. La note moyenne attribuée aux copies dans le groupe des correcteurs « sévères » fut égale à 12,4 alors qu'elle était égale à 11,6 pour les soi-disant « indulgents ».

La Figure 1 fournit un autre témoignage des variations inter individuelles ; elle résume les résultats d'une étude diffusée en 1969 par le C.R.D.P. de Lyon⁶, qui fit corriger 3 copies d'élèves ayant à résoudre le même problème à 150 professeurs de mathématiques exerçant dans une classe de troisième.

³ Enquête réalisée entre 1931 et la seconde guerre mondiale et impulsée par l'International Institute of Education, de Colombia, aux USA. Il s'agit d'une étude approfondie sur les examens officiels de plusieurs pays, fonctionnant via des commissions nationales : américaine, anglaise, écossaise, finlandaise, française, suisse et allemande (cette dernière étant arrêtée par la nazification de 1933).

⁴ Se reporter à : Bruno Suchaut. La loterie des notes au bac : un réexamen de l'arbitraire de la notation des élèves. Document de travail de l'IREDU 2008-03. 2008. <halshs-00260958v1>, pour une étude récente sur les défauts de fidélité dans la Correction de l'épreuve de Sciences Economiques et Sociale du baccalauréat.

⁵ <http://www.irdp.ch/edumetrie/lexique/spearman.htm>

⁶ C.R.D.P., *Docimologie et Examens*, Lyon, I.P.N, 1969

Max. 20	Moyenne	Marge de variation
Elève I	5,70	0,5 - 11,5
Elève II	16	11,5 - 20
Elève III	8	3,5 - 11,5

Figure 1: Résultats de correction de 3 copies de mathématiques par 150 correcteurs.
Figure extraite de De Landsheere, G. (1980).

Sensibilité de l'évaluation en mathématiques

L'enquête Carnegie mit en évidence que les matières les plus normées, comme les mathématiques, la physique ou encore le latin donnent lieu à des notes plus diverses (Figure 2). Pour ces matières « normées », l'échelle à 21 degrés est utilisée sur quasiment toute, voire toute son étendue, et par conséquent, le recours à cette échelle paraît cohérent, ou du moins, plus adapté que dans certaines disciplines comme la philosophie ou le français pour lesquelles les distributions se caractérisent par une forte « concentration » des notes.

	Distributions concentrées	Distributions étendues
Anglais	2,86	3,92
Mathématiques	4,00	4,30
Composition française	2,91	3,85
Physique	3,31	3,63
Version latine	3,70	4,66
Philosophie	2,22	4,25

Figure 2: Ecart-types observés sur des séries de notes dans différentes matières.
Figure extraite d'Abenot, Y. (1996)

Validité de l'évaluation en mathématiques

La note attribuée à une production ne reflète et ne mesure pas toujours exclusivement la valeur intrinsèque de celle-ci. Les effets influant la validité d'une évaluation peuvent se classer selon 4 principaux groupes : effets de contamination, effets de stéréotypie, effet de halo et effets parasites.

Les effets de contamination traduisent le fait que l'avis des confrères, par le biais notamment du dossier scolaire de l'élève, a une influence non négligeable dans le processus de correction d'une copie. Selon Caverni, J.-P. et al. (1975), cette contamination serait d'autant plus importante que le niveau de la copie est élevé.

Les effets de stéréotypie traduisent *une immuabilité plus ou moins accusée qui s'installe dans le jugement porté sur l'élève*⁷. L'enseignant a tendance à « enfermer l'élève » dans une position figée dépendant des premières notes obtenues par celui-ci en début d'année scolaire.

L'effet de halo illustre le fait que le jugement que l'on se fait d'un élève est influencé par des critères tels que la tenue vestimentaire, la sympathie transmise par cet élève, l'attitude face à l'école, la manière de s'exprimer (verbalisation, diction), la présentation d'une copie ou d'un élève, le graphisme d'une écriture, la profession des parents, entre autres exemples. R.

⁷ Citation extraite de De Landsheere, G. (1980)

Weiss (1969)⁸ examina les notes d'une série de compositions d'arithmétique attribuées sous un préjugé favorable et défavorable, respectivement (Figure 3).

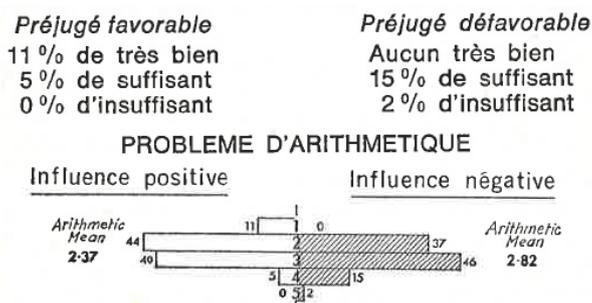


Figure 3: Mise en évidence par R. Weiss de l'effet de Halo, à l'occasion de la correction d'une production d'arithmétique.

Figure extraite de De Landsheere, G. (1980).

Parmi les effets parasites Abernot, Y. (1996) cite :

- l'état d'esprit de l'enseignant : e.g. favoritisme face à la copie d'un élève paraissant « méritant » ;
- la fatigue/charge de travail : quantité de copies à corriger, période de l'année scolaire ;
- l'image que l'enseignant peut avoir de lui-même et donner aux autres : un enseignant s'éloignant du patron statistique gaussien classique : « peu de très mauvaises notes, peu d'excellentes notes, beaucoup de notes moyennes » pourrait craindre d'être « pointé du doigt », aussi bien par ses élèves que par ses collègues ;
- la première impression donnée par une copie à un enseignant (Noizet, G. Caverni J.P. 1978) ;
- les effets d'ordre et de contraste.

De plus, selon Pierre Merle (1998): « *le correcteur établit au début de sa correction un ensemble d'exigences à partir de la lecture des premières copies et ces exigences servent de référence dans la suite de son travail de correction.* »

Dans le cadre de la présente recherche, on se propose de distinguer clairement les effets d'ordre au sens strict, des effets dits de contraste, ces derniers s'inscrivant naturellement dans la catégorie des effets d'ordre. Par effet d'ordre au sens strict, on entend l'influence du rang de traitement d'une copie lors de la correction d'une série de copies. Bonniol J.J. (1972), dans le cadre de sa thèse de doctorat, examina de façon approfondie les effets d'ordre (sens strict) et de contraste.

Concernant les effets d'ordre au sens strict, selon Bonniol J.J. (1972), les correcteurs seraient plus sévères en fin de correction plutôt qu'au début, ce qui se traduit statistiquement par une diminution de la moyenne des notes attribuées, au fur et à mesure du processus de correction d'une série de copies (les résultats ayant mené à cette thèse sont présentés en Figure 2 dans le paragraphe II.1 dédié au protocole expérimental). Par ailleurs, Betz, D. (1974) constatât que la dispersion des notes d'un évaluateur augmente au cours du processus de correction d'une pile de copies, ce qui se traduit statistiquement par une augmentation de l'écart type (ou encore de l'étendue) des notes attribuées au fur et à mesure du processus de correction d'une série de copies. Sur la base de ces constatations,

⁸ R. Weiss, The Reliability of the Number Marking System: An Austrian Study. In J. Lauwerys et D. Scalon, *Examinations*, Londres, Evans Brothers, 1969, pp. 101-107.

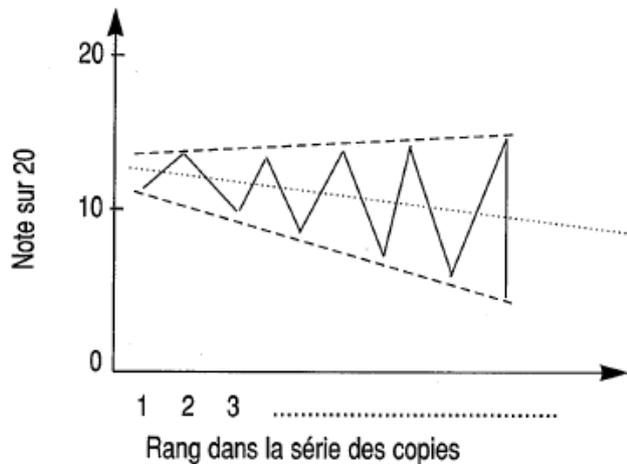
Abernot Y. (1996), résume dans la Figure 4 les observations ayant attrait aux effets d'ordre (au sens strict du terme).

Figure 4: Les effets d'ordre (sens strict) dans le processus de correction d'une série de copies.

Graphique extrait de l'ouvrage de Y. Abernot (1996). La ligne en pointillés fins de pente négative reflète la diminution de la moyenne des notes attribuées.

Les deux lignes en pointillés larges illustrent le fait que la distribution des notes augmente au fur et à mesure de la correction d'une série de copies.

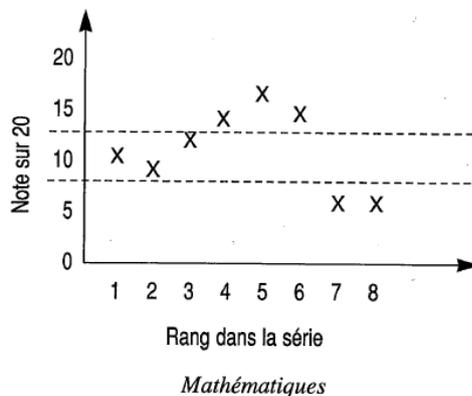
La ligne brisée n'est qu'une aide à la visualisation du phénomène général.



Pour ce qui est des effets de contraste, toujours selon Bonniol J.J, (1972), la note d'une copie dépendrait en partie de la ou les copies précédentes. En effet, pour expliquer et illustrer le phénomène, Bonniol, J.J et Piolat, M. (1971) placèrent dans une série de copies d'anglais et de mathématiques soit de très bonnes ou soit de très mauvaises copies, désignées « ancrs hautes » et « ancrs basses », respectivement, afin d'examiner leur effet sur les copies suivantes : pour certaines disciplines, l'effet de contraste serait important surtout après les ancrs basses, alors que pour d'autres, comme c'est le cas des mathématiques, ce seraient davantage les ancrs hautes et lourdes (plusieurs copies excellentes) qui entraîneraient la sous-estimation des copies suivantes (Figure 5). De plus, ces phénomènes seraient accentués lorsque les ancrs se trouvent en début de série.

Figure 5: Effets de contraste dans la correction d'une série de copies, en mathématiques.

Graphique extrait de l'ouvrage de Y. Abernot (1996). L'effet des copies de rang 4, 5^e et 6 (ancres « hautes ») se répercute sur les copies de rang 7 et 8.



Il est à noter que Noizet, G. et Caverni, J.-P., dans leur ouvrage datant de 1978, reportent un autre effet de contraste, également mis en exergue par les expériences de Bonniol, J.J. (1965), à savoir, le phénomène de sous-évaluation de la première copie. Ce phénomène, qui paraît aller à l'encontre des effets de contraste susmentionnés, trouverait son explication dans le fait que lorsqu'un correcteur débute la correction d'une pile de copies, la seule référence ou élément de comparaison dont il dispose est le produit-norme, c'est-à-dire la correction exacte de l'évaluation. Relativement à ce produit-norme qui revêt un caractère « idéal » aux yeux du correcteur, la première copie contient forcément des erreurs, et se retrouve donc, par suite, sous-évaluée, le produit-norme agissant alors comme ancre haute.

5. Questions de recherche

Cette recherche, inscrite dans un cadre docimologique, ne traite pas des questions de sensibilité (variations intra-individuelle ou inter-individuelle) ou de fidélité (adéquation du système de mesure utilisé avec la production évaluée). En effet, elle se centre exclusivement sur une question de validité de l'évaluation, à savoir, les effets d'ordre, et ce, au sens strict, c'est-à-dire l'influence du rang de traitement d'une copie lors de la correction d'une série de copie.

La problématique de recherche est déclinée sous forme de plusieurs interrogations :

- ✓ **Question 1** : Est-il possible, dans le cadre de l'évaluation en mathématiques, de mettre en évidence des effets d'ordre (au sens strict) lors de la correction d'une série de copies ?
- ✓ **Question 2** : Le cas échéant (cf. question 1), peut-on réussir à caractériser davantage les tendances mises en évidence par Bonniol, J.J. (1972) et Betz, D. (1974), (Figure 4) ?
- ✓ **Question 3** : Le cas échéant (cf. question 1), l'expérience des correcteurs, en termes d'années d'enseignement, a-t-elle une influence sur les effets d'ordre ?

Afin de répondre à ces questions, trois hypothèses ont été formulées afin d'orienter le traitement des données :

- ✓ **Hypothèse H₁** : au fur et à mesure de la correction la moyenne diminue et l'écart type augmente ;
- ✓ **Hypothèse H₂** : au fur et à mesure de la correction la moyenne diminue ;
- ✓ **Hypothèse H₃** : au fur et à mesure de la correction l'écart type augmente ;

II. Méthodologie de recueil et traitement des données

1. Protocole expérimental : type de données recueillies

Le protocole expérimental employé dans les études antérieures par Bonniol J.J., (1972) et Bonniol, J.J et Piolat, M., (1971), n'a pu être reconduit dans le cadre de la présente recherche. En effet, ces auteurs procédèrent à faire corriger une même série de copies – composée de 26 versions d'anglais d'une classe de terminale - par 18 correcteurs différents, et ce, dans l'ordre direct pour 9 d'entre eux (copie 1 à 26), et dans l'ordre inverse pour les 9 autres (copie 26 à 1).

	Ordre direct (9 évaluateurs)	Ordre inverse (9 évaluateurs)
Première moitié	10,75	9,85
Deuxième moitié	9,24	10,14
Premier tiers	11,17	9,01
Troisième tiers	9,18	10,36

Tableau 2 : Moyenne des notes attribuées à des versions anglaises, en fonction du sous-paquet considéré, selon l'ordre d'évaluation. Données d'étude de Bonniol, J.J. (1965).

Données reprises et mises en forme par Noizet, G., & Caverni, J.-P. (1978) d'après les données de Bonniol, J.J. (1965) : pour chacune des 18 séries de notes, la moyenne a été ramenée à 10 (centrage) et l'écart type ramené à 3 (réduction).

Ce protocole expérimental, nécessitant de convoquer une vingtaine de correcteurs un même jour, afin de procéder à la correction d'une même pile de copies dans des conditions « environnementales » identiques (heure, nombre de copies à corriger, etc.), ne pouvait raisonnablement être mis en place de nouveau. Par conséquent, le protocole expérimental adopté dans la présente recherche consiste à recueillir des données de correction « réelles », c'est-à-dire des séries de notes attribuées par des enseignants de mathématiques volontaires, à l'occasion de la correction de piles de copies d'une ou de plusieurs des classes dont ils sont les enseignants.

Rappelons finalement que les données expérimentales analysées plus bas sont donc des données issues de situation de correction « réelles ». Aussi, les données ont-elles pu être impactées à la fois par des effets parasites tels que les effets d'ordre, mais également par d'autres effets influant sur la validité d'une évaluation, comme par exemple, ceux de contamination, de stéréotypie et de halo. Il s'agit donc là d'un biais découlant directement du protocole expérimental choisi, qu'il est important de prendre en considération dans les phases d'analyse et de discussion postérieures (cf. paragraphes **Erreur ! Source du renvoi introuvable.** et **Erreur ! Source du renvoi introuvable.**).

2. Protocole expérimental : recueil des données

Les données ont été recueillies au moyen d'un questionnaire destiné à l'ensemble des enseignants de mathématiques de l'EA. Le questionnaire, consultable en Annexe 1, s'accompagne d'une lettre explicative et d'un mode d'emploi.

Afin d'éviter l'introduction d'un biais supplémentaire dans le protocole expérimental, l'objet de la recherche n'a pas été explicité avant la fin de la période de recueil des données. Les correcteurs ayant participé à la collecte de données sont donc des « correcteurs naïfs », c'est à dire non informés des objectifs de l'expérience.

Le questionnaire a été soumis par mail le 19 octobre 2015, pendant les vacances de la Toussaint, de manière à ce que les enseignants sollicités aient le temps de s'imprégner de la marche à suivre pour collaborer à cette collecte de données, et, le cas échéant, poser les questions qui leur semblaient pertinentes. La période de recueil de données s'est étalée du 19 octobre au 31 décembre 2015.

En plus des données de correction à strictement parler, les correcteurs participant à la collecte ont dû systématiquement indiquer :

- ✓ leur expérience en termes d'années d'enseignement ;
- ✓ si la correction a été faite en une seule fois ou en plusieurs fois ;
- ✓ si la correction a été faite « copie par copie » ou « exercice par exercice ».

Au total, 30 enseignants de mathématiques de l'EA ont participé à la collecte de données ; 96 formulaires ont été reçus, dont 33 issus d'une correction de type « exercice par exercice » et 63 issus d'une correction de type « copie par copie ».

3. Protocole expérimental : traitement des données

Un nombre compris entre 1 et 30 a été attribué à chacun des 30 correcteurs de manière à traiter les données dans le strict respect de l'anonymat.

Au total, sur les 96 formulaires reçus, 2 formulaires correspondant à une correction du type « copie par copie » n'ont pas été pris en compte dans la phase de traitement des données pour les raisons suivantes :

- ✓ un formulaire, le seul issu du correcteur 12, ne fait apparaître que trop peu de notes (série de 6 notes) ; et
- ✓ l'autre formulaire, issu du correcteur 3, a été renseigné de manière incohérente (erreur de remplissage).

Les 94 séries de copies ayant finalement été analysées sont détaillées dans le Tableau 3 en fonction du correcteur dont elles sont issues et du type de correction qui leur correspond. Notons que certains correcteurs corrigent tantôt « exercice par exercice », tantôt « copie par copie » (cas des correcteurs 4 et 11).

	Correction de type exercice par exercice	Correction de type copie par copie
	1: 1 2: 3 4: 2 5: 1 7: 2 8: 1 11: 5 18: 4 19: 2 20: 2 23: 2 24: 1 25: 1 27: 5 30: 1	3: 1 4: 6 6: 1 9: 3 10: 12 11: 6 13: 3 14: 3 15: 1 16: 3 17: 1 21: 4 22: 12 26: 2 28: 2 29: 1
Nombre total de séries analysées: 94	33	61

Tableau 3 : Provenance (correcteur) des séries de copies prises en compte dans la phase de traitement des données en fonction du type de correction leur correspondant.

Dans chaque étiquette, l'identifiant du correcteur apparaît en gras et le nombre de séries lui correspondant est indiqué en italique.

Après une phase de prétraitement des données, trois méthodes de traitement ont été adoptées et sont décrites dans le Tableau 4.

Prétraitement des données

Afin de comparer aisément les données issues de différents correcteurs, un centrage et une réduction des notes ont été réalisés, dans le but de ne conserver que les écarts relatifs entre les notes attribuées aux copies d'une même série (cf. paragraphe I.3. Bref rappel de notions statistiques).

Dans ce sens, pour chaque série, les notes ont été modifiées afin de ramener la moyenne de la série à 10 (opération de « centrage ») et l'écart type de la série à 3 (opération de « réduction »). Les valeurs 10 et 3, respectivement pour la moyenne après centrage et l'écart type après réduction, ont été choisies afin de conserver l'esprit de l'analyse des données de Bonniol, J.J. (1965) conduite par Noizet, G., et Caverni, J.-P. en 1978, bien que les protocoles expérimentaux diffèrent (cf. paragraphe II.1. Protocole expérimental : type de données recueillies et Tableau 2).

Traitement des données : méthode 1, méthode 2 et méthode 3

Méthode 1 : Méthode des moyennes et écarts types par sous-ensemble	Méthode 2 : méthode des moyennes et écarts types glissants.	Méthode 3 : méthode des moyennes et écarts types par rang.
<p>Cette méthode repose sur le fait de séparer chaque série de notes en sous-ensembles de taille égale (ou quasiment égale), comme par exemple, première moitié/deuxième moitié ou premier tiers/deuxième tiers/troisième tiers, ou premier quart/second quart/troisième quart/quatrième quart. Ensuite, sont calculés la moyenne et l'écart type des notes dans chaque sous-ensemble, afin d'examiner de potentielles variations de celles-ci en fonction du sous-ensemble considéré.</p> <p>Cette méthode a été utilisée par Noizet, G., & Caverni, J.-P., en 1978 ; les résultats obtenus par ces derniers sont consignés dans le Tableau 2.</p>	<p>Soit N le nombre de notes que comporte une série de donnée et k un nombre fixé entre 2 et N. Cette méthode consiste à calculer la moyenne et l'écart type de k notes successives, en observant, entre chaque calcul, un décalage de 1 rang. Le premier calcul s'effectue sur les notes de rang 1 à k, le second calcul sur les notes de rang 2 à $k + 1$, ainsi de suite jusqu'au dernier calcul qui s'opère sur les notes de rang $N - k + 1$ à N. Cette méthode fournit alors $N - k + 1$ couples de valeurs moyenne/écart type.</p>	<p>Cette méthode consiste à calculer la moyenne et l'écart type des notes rang par rang, c'est-à-dire, toutes séries de notes confondues. On calcule la moyenne et l'écart type en utilisant la première note de toutes les séries de notes, puis la moyenne et l'écart type en utilisant la note de rang 2 de toutes les séries de notes, ainsi de suite. La limitation de cette méthode réside dans le fait que les séries de notes ne sont pas toutes de même taille, étant donné qu'elles proviennent de classes différentes et d'évaluations différentes.</p>

Tableau 4 : Méthodes de traitement des données.

4. Protocole expérimental : analyse des données

Ce paragraphe s'attache à décrire la méthode d'analyse des données adoptée dans le cadre de ce travail de recherche.

Dans le cas d'une correction de type « exercice par exercice », le paquet de copie est « inversé » dès lors que le correcteur entame la correction d'un nouvel exercice. De ce fait, il paraît cohérent de penser que les éventuelles tendances observées pour les paramètres moyenne et écart type en fonction du rang de traitement des copies ne sont que le fruit du hasard.

Par suite, étant donné que dans le cas d'une correction de type « exercice par exercice », les tendances observées peuvent être interprétés comme étant le fruit du hasard, le principe de l'analyse des données adopté dans le cadre de ce travail de recherche repose sur le fait de comparer l'occurrence des manifestations en accord avec les hypothèses de travail H_1 , H_2 et H_3 entre une correction de type « exercice par exercice » et une correction de type « copie par copie ».

Le principe de l'analyse des données est détaillé ci-dessous dans de plus amples détails avec l'hypothèse H_1 . Il est identique pour les deux autres hypothèses de travail (H_2 et H_3).

L'idée est de comparer la proportion p_{exo} de séries (de copies) corroborant l'hypothèse H_1 dans le cas d'une correction de type « exercice par exercice » avec la proportion p_{copie} de séries corroborant la même hypothèse dans le cas d'une correction de type « copie par copie ». Si la proportion p_{copie} est largement supérieure à la proportion p_{exo} , l'hypothèse H_1 pourrait alors être confirmée.

III. Analyse réflexive des résultats de la recherche

Remarque importante: Les valeurs de moyenne et d'écart type représentées dans tous les graphiques de la partie 0, se lisent sur l'axe de gauche et de droite, respectivement.

1. Méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en quarts.

Après séparation des séries de notes en quarts, la moyenne et l'écart type ont été calculés dans chacun des 4 sous-paquets (1^{er} quart, 2^e quart, 3^e quart, 4^e quart). Ensuite, pour chaque série de note, l'évolution des moyennes et l'évolution des écarts types ont été représentées dans un graphique à deux axes.

Les planches de résultats présentant les graphiques obtenus par la méthode 1 de traitement des données (séparation des séries en quart) sont placées en Annexe 2.

Afin d'analyser les graphiques obtenus, 5 profils « types » ont été définis (Figure 6).

Profil	Exemple	Description du profil
A	<p>Graphique A: Moyenne diminue et écart type augmente. Correcteur 3.</p>	La moyenne diminue et l'écart type augmente. Ce profil illustre l'hypothèse H ₁ .
B	<p>Graphique B: Moyenne diminue et aucune tendance n'est observée pour l'écart type ou l'écart type diminue. Correcteur 10.</p>	La moyenne diminue et aucune tendance n'est observée pour l'écart type ou l'écart type diminue. Ce profil illustre l'hypothèse H ₂ .
C	<p>Graphique C: L'écart type augmente et aucune tendance n'est observée pour la moyenne ou la moyenne augmente. Correcteur 27.</p>	L'écart type augmente et aucune tendance n'est observée pour la moyenne ou la moyenne augmente. Ce profil illustre l'hypothèse H ₃ .
D	<p>Graphique D: La moyenne augmente et l'écart type diminue. Correcteur 21.</p>	La moyenne augmente et l'écart type diminue. Ce profil met en défaut l'hypothèse H ₁ .
E	<p>Graphique E: Profil ne laissant apparaître aucune tendance remarquable : profil en dents de scies, ou « en miroir » (lorsque la moyenne augmente, l'écart type diminue et vice-versa). Correcteur 7.</p>	Profil ne laissant apparaître aucune tendance remarquable : profil en dents de scies, ou « en miroir » (lorsque la moyenne augmente, l'écart type diminue et vice-versa).

Figure 6 : Profils « types » définis dans le cadre de l'analyse des données traitées par la méthode 1 dans le cas où les séries ont été séparées en quarts.

En plus des 5 profils types définis plus haut, une « règle de tri » a été adoptée de manière à « trancher » quand à la classification de certains graphiques laissant apparaître une tendance à la diminution pour la moyenne. En effet si la moyenne montre une tendance à la diminution mais que ce paramètre est constant entre deux sous-paquets consécutifs ou que ce paramètre augmente de moins de 10% par rapport à la diminution totale enregistrée entre le premier et le quatrième sous-paquet, il a été considéré que la moyenne diminuait. Par suite, ce graphique a été classifié sous le profil A ou B en fonction du comportement de l'écart type ;

Les résultats d'analyses des graphiques correspondant à une correction de type « exercice par exercice » et de type « copie par copie » sont consignés dans le Tableau 5 et le Tableau 6, respectivement.

Correction de type « exercice par exercice »	Profil	Nombre de séries de copies présentant ce profil	Correcteurs et nombre de séries de copie présentant ce profil
	A	1 soit 3,0 %	2: 1
	B	3 soit 9,1 %	2: 1 4: 1 27: 1
	C	2 soit 6,1 %	11: 1 27: 1
	D	0 soit 0,0 %	
	E	27 soit 81,8 %	1: 1 2: 1 4: 1 5: 1 7: 2 8: 1 11: 4 18: 4 19: 2 20: 2 23: 2 24: 1 25: 1 27: 3 30: 1
		Total : 33 séries (100 %)	

Tableau 5 : Résultats d'analyses des graphiques correspondant à une correction de type "exercice par exercice" (méthode 1 - séparation des séries en quarts).

Correction de type « copie par copie »	Profil	Nombre de séries de copies présentant ce profil	Correcteurs et nombre de séries de copie présentant ce profil
	A	1 soit 1,6 %	3: 1
	B	3 soit 4,9 %	9: 1 10: 1 13: 1
	C	2 soit 3,3 %	4: 1 22: 1
	D	1 soit 1,6 %	21: 1
	E	54 soit 88,5 %	4: 5 6: 1 9: 2 10: 11 11: 6 13: 2 14: 3 15: 1 16: 3 17: 1 21: 3 22: 11 26: 2 28: 2 29: 1
		Total : 61 séries (100 %)	

Tableau 6 : Résultats d'analyses des graphiques correspondant à une correction de type "copie par copie" (méthode 1 - séparation des séries en quarts).

Dans chaque étiquette, l'identifiant du correcteur apparaît en gras et le nombre de séries lui correspondant est indiqué en italique.

Le Tableau 7 confronte les résultats obtenus en fonction du type de correction employée.

Profil	Correction de type « exercice par exercice »	Correction de type « copie par copie »
A	3,0 %	1,6 %
B	9,1 %	4,9 %
C	6,1 %	3,3 %
D	0,0 %	1,6 %
E	81,8 %	88,5 %

Tableau 7 : Confrontation des résultats obtenus en fonction du type de correction (méthode 1 - séparation des séries en quarts).

Le traitement des données par la méthode 1 dite des moyennes et écarts types par sous-ensembles ne permet pas de confirmer les hypothèses H_1 , H_2 et H_3 , dans le cas où les séries de notes sont séparées en quarts.

De manière générale, le profil A, illustrant l'hypothèse H_1 , n'est pas observé davantage dans le cas d'une correction de type « copie par copie » que dans le cas d'une correction de type « exercice par exercice ». Il en va de même pour les profils B et C illustrant respectivement les hypothèses H_2 et H_3 .

De manière plus précise, les profils A, B et C sont moins observés dans le cas d'une correction de type « copie par copie » que dans le cas d'une correction de type « exercice par exercice », et le profil D est observé une fois dans le cas d'une correction de type « copie par copie » alors qu'il n'a pas été observé dans le cas d'une correction de type « exercice par exercice ». Ces deux dernières remarques pourraient amener à infirmer les hypothèses de travail (H_1 , H_2 et H_3). Cependant, remarquons d'une part que les pourcentages d'observation des différents profils ne sont pas établis identiquement pour les deux types de correction : l'analyse se base sur l'étude de 33 séries correspondant à une correction de type « exercice par exercice » et 61 séries correspondant à une correction de type « copie par copie ». D'autre part, il se peut que la séparation des séries en quart engendre une prise en compte trop importante d'autres facteurs d'influence comme par exemple les effets de contraste ou simplement l'apparition d'une note extrême dans la série.

2. Méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en moitiés.

Après séparation des séries de notes en moitiés, la moyenne et l'écart type ont été calculés dans chacun des 2 sous-paquets (1^{re} moitié, 2^e moitié). Ensuite, pour chaque série de note, l'évolution des moyennes et l'évolution des écarts types ont été représentées dans un graphique à deux axes.

Les planches de résultats présentant les graphiques obtenus par la méthode 1 de traitement des données (séparation des séries en moitiés) sont placées en Annexe 3.

Afin d'analyser les graphiques obtenus, 4 profils « types » ont été définis (Figure 7).

Profil	Exemple	Description du profil
A	<p>Graphique pour le profil A (Correcteur 1). L'axe des ordonnées gauche (0-12) mesure l'écart type, et l'axe des ordonnées droite (0-6) mesure la moyenne. La moyenne (ligne pointillée) diminue de 10 à 8, et l'écart type (ligne en tirets) augmente de 2 à 3.</p>	La moyenne diminue et l'écart type augmente. Ce profil illustre l'hypothèse H ₁ .
B	<p>Graphique pour le profil B (Correcteur 2). L'axe des ordonnées gauche (0-12) mesure l'écart type, et l'axe des ordonnées droite (0-6) mesure la moyenne. La moyenne (ligne pointillée) diminue de 10 à 8, et l'écart type (ligne en tirets) diminue de 4 à 3.</p>	La moyenne diminue et l'écart type diminue ou est constant. Ce profil illustre l'hypothèse H ₂ .
C	<p>Graphique pour le profil C (Correcteur 5). L'axe des ordonnées gauche (0-12) mesure l'écart type, et l'axe des ordonnées droite (0-6) mesure la moyenne. La moyenne (ligne pointillée) augmente de 6 à 7, et l'écart type (ligne en tirets) augmente de 4 à 5.</p>	L'écart type augmente et la moyenne augmente ou est constante. Ce profil illustre l'hypothèse H ₃ .
D	<p>Graphique pour le profil D (Correcteur 25). L'axe des ordonnées gauche (0-12) mesure l'écart type, et l'axe des ordonnées droite (0-6) mesure la moyenne. La moyenne (ligne pointillée) augmente de 7 à 10, et l'écart type (ligne en tirets) diminue de 3 à 2.</p>	La moyenne augmente et l'écart type diminue. Ce profil met en défaut l'hypothèse H ₁ .

Figure 7 : Profils « types » définis dans le cadre de l'analyse des données traitées par la méthode 1 dans le cas où les séries ont été séparées en moitiés.

Les résultats d'analyses des graphiques correspondant à une correction de type « exercice par exercice » et de type « copie par copie » sont consignés dans le Tableau 8 et le Tableau 9, respectivement.

Le Tableau 10 confronte finalement les résultats obtenus en fonction du type de correction employée. Le traitement des données par la méthode 1 dite des moyennes et écarts types par sous-ensembles, lorsque les séries de notes sont séparées en moitiés, aboutit aux observations suivantes :

- le profil A, illustrant l'hypothèse H₁, est davantage observé (+5,3 %) dans le cadre d'une correction de type « copie par copie » que dans le cadre d'une correction de type « exercice par exercice » (29,5 % des séries répondent au profil A dans le cadre

d'une correction de type « copie par copie » contre 24,2 % dans le cadre d'une correction de type « exercice par exercice » ;

- les profils B et C, illustrant respectivement les hypothèses H_2 et H_3 , sont un peu moins observés (-2,4 % et -1,2 %, respectivement) dans le cadre d'une correction de type « copie par copie » que dans le cadre d'une correction de type « exercice par exercice » ;
- le profil D, mettant en défaut l'hypothèse H_1 , est moins observé (-1,5 %) dans le cadre d'une correction de type « copie par copie » que dans le cadre d'une correction de type « exercice par exercice ».

Les trois observations décrites plus haut pourraient amener à confirmer l'hypothèse de travail H_1 selon laquelle la moyenne diminue et l'écart type augmente au fur et à mesure du processus de correction, et ce, dans le cadre d'une correction de type « copie par copie ».

Correction de type « exercice par exercice »	Profil	Nombre de séries de copies présentant ce profil	Correcteurs et nombre de séries de copie présentant ce profil
	A	8 soit 24,2 %	1: 1 2: 1 7: 1 11: 1 18: 2 27: 1 30: 1
	B	10 soit 30,3 %	2: 1 4: 2 11: 2 20: 2 23: 1 24: 1 27: 1
	C	8 soit 24,2 %	2: 1 5: 1 11: 1 18: 2 19: 1 27: 2
	D	7 soit 21,2 %	7: 1 8: 1 11: 1 19: 1 23: 1 25: 1 27: 1
	Total : 33 séries (100 %)		

Tableau 8 : Résultats d'analyses des graphiques correspondant à une correction de type "exercice par exercice" (méthode 1 - séparation des séries en moitiés).

Dans chaque étiquette, l'identifiant du correcteur apparaît en gras et le nombre de séries lui correspondant est indiqué en italique.

Correction de type « copie par copie »	Profil	Nombre de séries de copies présentant ce profil	Correcteurs et nombre de séries de copie présentant ce profil
	A	18 soit 29,5 %	3: 1 4: 2 6: 1 9: 1 10: 2 11: 1 13: 2 14: 1 15: 1 21: 1 22: 5
	B	17 soit 27,9 %	4: 3 9: 2 10: 3 11: 3 14: 1 16: 1 17: 1 22: 3
	C	14 soit 23,0 %	10: 4 14: 1 16: 2 21: 1 22: 1 26: 2 28: 2 29: 1
	D	12 soit 19,7 %	4: 1 10: 3 11: 2 13: 1 21: 2 22: 3
	Total : 61 séries (100 %)		

Tableau 9 : Résultats d'analyses des graphiques correspondant à une correction de type "copie par copie" (méthode 1 - séparation des séries en moitiés).

Dans chaque étiquette, l'identifiant du correcteur apparaît en gras et le nombre de séries lui correspondant est indiqué en italique.

Profil	Correction de type « exercice par exercice »	Correction de type « copie par copie »
A	24,2 %	29,5 %
B	30,3 %	27,9 %
C	24,2 %	23,0 %
D	21,2 %	19,7 %

Tableau 10 : Confrontation des résultats obtenus en fonction du type de correction (méthode 1 - séparation des séries en moitiés).

3. Méthode 2 : moyennes et écarts types glissants.

La méthode 2 dite des moyennes et écarts types glissants nécessite le choix d'un réel k compris entre 2 et N afin de calculer les paramètres d'intérêt (moyenne et écart type) sur k notes successives, et ce, en observant, entre chaque calcul, un décalage de 1 rang (cf. paragraphe II.3).

Dans le cas de la présente étude, la méthode 2 a d'abord été mise en place pour deux valeurs de k différentes, 3 et 6, respectivement, afin de choisir celle de ces deux valeurs aboutissant aux résultats les plus pertinents par rapport à l'analyse des données postérieure.

Le « risque » accompagnant le choix d'une petite valeur de k est de prendre en compte excessivement les notes extrême, le cas échéant. D'autre part, le « risque » encouru lors de l'utilisation d'une trop grande valeur de k est « un lissage » excessif des paramètres obtenus autour des valeurs de ceux-ci calculés sur la série entière, ne laissant par conséquent plus apparaître de tendance.

La Figure 8 illustre les résultats obtenus pour une série de données en particulier (issue du correcteur 3), et ce pour $k=3$ et $k=6$, respectivement. On observe comment la valeur de k influe sur l'aspect plus ou moins lissé du graphique. Afin de faciliter la lecture des résultats obtenus par cette méthode de traitement des données, il a été décidé de travailler avec une valeur de $k=6$.

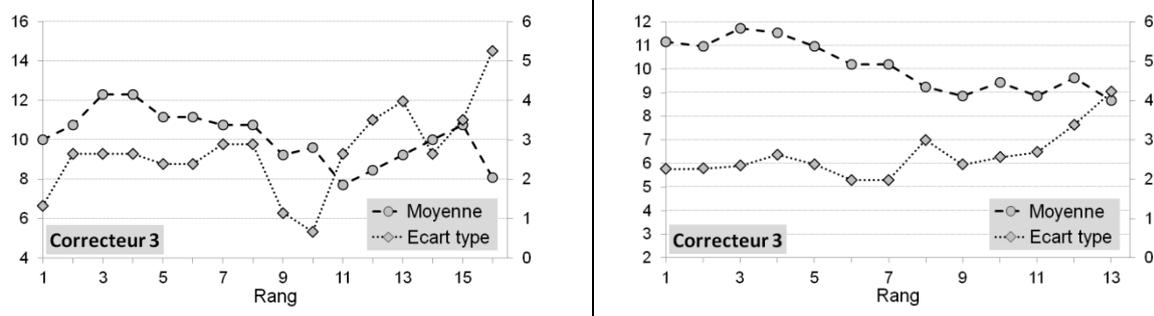


Figure 8 : Exemple d'une série de données traitée par la méthode 2 avec $k=3$ (à gauche) et $k=6$ (à droite).

En abscisse, l'étiquette rang signifie non pas le rang d'une copie dans la série considérée mais le rang du calcul. En effet, la méthode 2 de traitement des données fournit $N - k + 1$ couples de valeur moyenne/écart type.

Les planches de résultats présentant les graphiques obtenus par la méthode 2 de traitement des données avec une valeur de $k=6$ sont placées en Annexe 4.

Bien que le choix « $k=6$ » permette un lissage des évolutions de la moyenne et de l'écart type, la lecture et la classification des graphiques reste globalement difficile du fait du niveau de détail important spécifique à cette méthode (selon la série de note considérée, le nombre de couples moyenne/écart type calculés est compris entre 6 et 31).

Par conséquent, en raison des difficultés évoquées plus haut, la méthode 2 de traitement des données n'a pas abouti sur une classification approfondie des graphiques selon différents types de profils illustrant respectivement les hypothèses de travail H_1 , H_2 et H_3 , à l'instar de l'analyse réalisée dans le cadre de la mise en place de la méthode 1 de traitement des données.

4. Méthode 3 : moyennes et écarts types par rang.

La méthode 3, dite des moyennes et écarts types par rang, consiste à calculer la moyenne et l'écart type des notes rang par rang, c'est-à-dire, toutes séries de notes confondues (cf. paragraphe II.3).

Sa mise en place dans le cadre de la présente recherche s'est vue entravée par le fait que les séries de notes considérées ne sont pas toutes de même taille : la taille des séries de notes considérées fluctue entre 11 et 32 copies dans le cas d'une correction de type « exercice par exercice » et entre 11 et 36 copies dans le cas d'une correction de type « copie par copie ». De ce fait, pour ne citer qu'un exemple, le calcul des paramètres (moyenne et écart type) au rang 11 se base sur la dernière copie des séries de taille 11 et la 11^e copie des séries de taille supérieure à 11. Dans les séries de taille supérieure à 11, la 11^e copie ne reflète pas nécessairement une « copie de fin de série », contrairement à la 11^e copie d'une série de taille 11.

Par ailleurs, les calculs ont été effectués tant que le nombre de copies était supérieur ou égal à 15. Ainsi, dans le cas d'une correction de type « exercice par exercice », les calculs de moyenne et d'écart type ont été réalisés jusqu'au rang 27, et dans le cas d'une correction de type « copie par copie », les calculs ont été réalisés jusqu'au rang 29.

Les graphiques obtenus par la méthode 3 de traitement des données sont présentés en Figure 9 et sont également placés en Annexe 5.

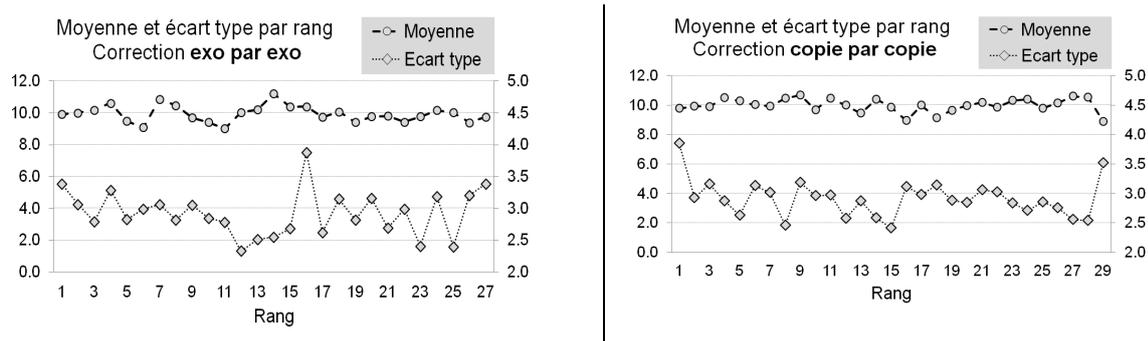


Figure 9: Graphiques obtenus par la méthode 3 de traitement des données.

La méthode 3 de traitement des données ne permet pas de confirmer les hypothèses de travail. En effet, les évolutions de la moyenne et de l'écart type obtenues dans le cas d'une correction ne laissent apparaître aucune tendance.

De plus, aucune différence n'est à relever entre les évolutions des paramètres obtenues dans le cas d'une correction de type « exercice par exercice » et « copie par copie ». Cette dernière observation pourrait nous amener à infirmer les hypothèses de travail (H_1 , H_2 et H_3). Néanmoins, au vu de la mise en place complexe de cette méthode de traitement des données, il semble pertinent de penser que cette méthode est de loin la méthode la moins adaptée au traitement des données dans le cadre de la présente recherche étant donné qu'il est nécessaire de gérer des séries de notes de taille très différentes.

Conclusion et perspectives d'exploitation professionnelle

La méthode 1 de traitement des données dite, *des moyennes et écarts types par sous-ensembles*, lorsque les séries sont séparées en moitiés, pourrait laisser penser à la validité de l'hypothèse de travail H_1 selon laquelle « au fur et à mesure de la correction, la moyenne diminue et l'écart type augmente ». Malgré cela, les effets d'ordres (au sens strict), tels qu'ils ont été décrits par Bonniol, J.J. (1972) et Betz, D. (1974), n'ont pas été mis en évidence de façon claire dans le cadre de ce projet de recherche restreint au cas de l'évaluation en mathématiques sur une échelle de 21 degrés. Par conséquent :

- il n'a pas été possible de proposer un modèle d'évolution plus précis des paramètres moyenne et écart type, au fur et à mesure de l'avancement du processus de correction d'une série de copies ;
- l'influence potentielle du facteur « expérience des correcteurs, en termes d'années d'enseignement », n'a pas pu être étudiée.

Bien que la principale hypothèse de travail (H_1) n'ait pu être confirmée, il serait néanmoins erroné de l'infirmier. En effet, les données expérimentales recueillies et analysées sont issues de corrections « réelles » ayant de ce fait pu être impactées à la fois par des effets parasites tels que les effets d'ordre, mais également par d'autres effets influant sur la validité d'une évaluation, comme par exemple, ceux de contamination, de stéréotypie et de halo.

Afin d'optimiser d'éventuelles investigations futures sur le sujet (étude simultanée de l'évolution des paramètres moyenne et écart type), il pourrait être judicieux d'adopter un protocole expérimental proche de celui mis en place par Bonniol J.J., (1972) et Bonniol, J.J et Piolat, M., (1971), et ce, en augmentant considérablement le nombre de correcteurs afin de s'assurer de la représentativité des données et de la validité des résultats d'analyse.

Pour conclure, ce travail de mémoire, et plus généralement, cette année en tant qu'enseignante stagiaire de mathématiques, ont été l'occasion de réfléchir et de porter un regard critique sur ma manière d'aborder une séance de classe, de corriger les productions de mes élèves, d'introduire telle ou telle notion aux élèves. Bien que les phénomènes supposés n'aient pas pu être mis en évidence dans le cadre de ce projet de recherche, rien ne prouve à l'heure actuelle que les effets d'ordre ne jouent pas un rôle important dans le processus de correction de type « copie par copie ». Ainsi, afin de les éviter, il peut être bon d'observer les quelques règles suivantes : reprendre les premières copies, commencer le processus de notation après avoir « parcouru » toute la série, adopter une correction du type « exercice par exercice ».

Les travaux en docimologie ne tendent pas nécessairement à assurer une homogénéisation du processus de notation, mais plutôt à une prise de conscience de la part des correcteurs des facteurs intervenant dans la notation. Par ailleurs, le but poursuivi n'est pas de démontrer que les professeurs notent « n'importe comment », mais au contraire d'analyser ce qui les influence, pour les rendre plus conscients de certains effets.

Les notes sont relatives, non seulement au groupe de référence, et à l'établissement scolaire, mais aussi à l'enseignant qui les distribue. Ceci n'a rien de scandaleux, mais il est indispensable d'en avoir conscience pour tirer le meilleur parti de l'évaluation. Il y a lieu de rendre les notes plus justes et plus utiles, non de les supprimer !

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

Bibliographie

- Abernot, Y. (1996). *Les Méthodes d'évaluation scolaire*. Paris : Dunod.
- Betz, D. (1974). Rhythmische Schwankungen als Fehler in der Notengebung. *Psychologie in Erziehung und Unterricht*, (21).
- Bonboir, A. (1972). *La docimologie*. Paris : Presses Universitaires de France.
- Bonniol, J.-J. (1972). *Les comportements d'estimation dans une tâche d'évaluation d'épreuves scolaires. Etude de quelques-uns de leurs déterminants*. (Thèse de 3ème cycle). Université de Provence, Aix-en-Provence.
- Bonniol, J.-J., & Piolat, M. (1971). Comparaison des effets d'ancrage obtenus dans une tâche d'évaluation. Expérience de multi-correction en mathématique et en anglais. In *Actes du XVIIe Congrès international de psychologie appliquée* (Vol. 8, p. 1179-1189).
- Bonniol, J.-J., & Vial, M. (1997). *Les modèles de l'évaluation: textes fondateurs avec commentaires*. Paris : De Boeck Univ.
- Caverni, J.-P. et al. (1975). Dépendances des évaluations scolaires par rapport à des évaluations antérieures, dans *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 Landsheere, G. (1980). *Evaluation continue et Examens. Précis de docimologie* (4^e éd.). Bruxelles-Paris : Labor-Nathan.
- Duquenne, R. (1967). *L'éducation Nationale*.
- Langouet, G., & Porlier, J.-C. (1994). *Mesure et Statistiques en milieu éducatif* (5^e éd.). Paris : ESF éditeur.
- Laugier, H., & Weinberg, D. (1936). *Elaboration statistique des données numériques de l'enquête sur la correction des épreuves du baccalauréat dans La correction des épreuves écrites dans les examens*. Paris : Maison du livre.
- Lauwerys, J., & Scalon, D. (1969). *Examinations*. Londres : Evans Brothers.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Noizet, G., & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Pieron, H. (1963). *Examens et docimologie*. Paris : Presses Universitaires de France.

Annexes

Annexe 1: Formulaire de recueil de données	32
Annexe 2: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en quarts.	35
Annexe 3: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en moitiés.	45
Annexe 4: Planches de résultats obtenus par la méthode 2 : moyennes et écarts types glissants.	55
Annexe 5: Résultats obtenus par la méthode 3 : moyennes et écarts types par rang.	65

Remarque importante: Les valeurs de moyenne et d'écart type représentées dans les graphiques des Annexes 2 à 5, se lisent sur l'axe de gauche et de droite, respectivement.

Annexe 1: Formulaire de recueil de données

Projet de mémoire – Formation ENFA – Année scolaire 2015-2016

CANCET Emmanuelle

Aux professeurs de mathématiques de l'enseignement secondaire

Objet : Sollicitation de votre aide pour le recueil de données expérimentales

Le 19 octobre 2015

Bonjour,

Dans le cadre de mon année de stage en alternance entre l'ENFA (Ecole Nationale de Formation Agronomique) et le Legta Toulouse-Auzeville, il m'est demandé de réaliser une recherche en sciences de l'éducation. J'ai décidé d'inscrire ce travail dans le cadre de l'évaluation en mathématiques sur une échelle de 21 degrés (de 0 à 20).

Afin de mener cette recherche, j'aurais besoin de recueillir des données « de terrain », directement issues de situations réelles, c'est à dire, directement issues de vos pratiques en tant qu'enseignant/e de mathématiques. Si vous êtes d'accord pour participer, vous trouverez en page 3 un formulaire de recueil de données, et en page 2, son mode d'emploi.

C'est à l'occasion de la correction d'une pile de copies que je vous demande de remplir le formulaire (un formulaire par pile de copies corrigée). Rien ne vous oblige à m'aider dans cette tâche, et votre participation n'implique pas un engagement sur la durée : vous pouvez collecter des données une seule fois (envoi d'un seul formulaire), ou « k » fois (envoi de « k » formulaires), pendant la période de collecte allant du 19 octobre 2015 au 31 décembre 2015.

Afin d'éviter tout biais dans le protocole expérimental, l'objet de la recherche ne peut, à ce stade, être explicité exhaustivement. Néanmoins, il semble important de préciser que cette étude ne vise en rien à vous « cataloguer » comme bon ou mauvais enseignant. De plus, les données seront traitées dans le respect de l'anonymat.

Enfin, si dans votre entourage, se trouvent des professeurs de mathématiques de collège ou de lycée – de l'enseignement général, technologie ou professionnel, de l'Education Nationale ou de l'Enseignement Agricole - susceptibles de participer à cette collecte de données, vous pouvez, bien entendu, leur transmettre ce questionnaire !

Par avance, merci pour votre aide,

Emmanuelle Cancet
Stagiaire mathématiques, enseignement agricole
emmanuelle.cancet@educagri.fr

p.s. Si vous souhaitez qu'au terme de la recherche, le compte-rendu (mémoire), vous soit adressé, merci de bien vouloir me le signaler par mail.

MODE D'EMPLOI DU FORMULAIRE

Le formulaire doit être rempli à l'occasion de la correction d'une pile de copies de la classe dont vous êtes l'enseignant (devoir maison ou sur table noté sur 20 points).

Période de recueil des données : du 19 octobre 2015 au 31 décembre 2015

Remise des formulaires : vous pouvez me retourner vos formulaires remplis sous format papier (dans mon casier au Legta Toulouse-Auzeville), ou bien au format numérique (scan ou traitement de texte ou pdf, etc.) par mail : emmanuelle.cancet@educagri.fr.

- 1) Indiquez la **date** à laquelle vous avez rempli le formulaire (en haut à gauche) et votre **identifiant correcteur** (en haut à droite).

Identifiant attribué par vous-même ; vous serez donc le seul à le connaître. Veillez simplement à ne pas le perdre et à utiliser toujours le même. Ce peut être un mot, un nombre, une expression...l'important est que vous vous en souveniez. Il ne doit refléter en rien votre personne.

- 2) Indiquez vos **années expérience** (en haut à droite).

Si vous n'avez pas d'expérience (vous êtes stagiaire ou c'est votre première rentrée), **indiquez 0**.

Si vous avez 1 an d'expérience (vous venez de faire votre 2nde rentrée), **indiquez 1**.

Si vous avez 2 ans d'expérience (vous venez de faire votre 3^{ème} rentrée), **indiquez 2**. etc...

- 3) En dessous de la date, entourez « oui » si vous avez corrigé le devoir « d'une traite », entourez « non » sinon. Si vous avez corrigé le devoir « d'une traite » mais que vous avez été amené à modifier certaines notes « ultérieurement » (quelques heures après, le lendemain, le surlendemain), entourez « oui ».

- 4) Dans le tableau, veuillez reporter les notes attribuées au contrôle que vous êtes en train de corriger, **en respectant scrupuleusement l'ordre de correction des copies première copie corrigée = rang de correction 1 ; seconde copie corrigée = rang de correction 2, etc.** Ne triez surtout pas les notes dans l'ordre croissant ou décroissant.

Dans l'éventualité où vous avez procédé à une rectification de certaines notes, en cours ou à l'issue de la correction de la pile de copie, merci d'utiliser la colonne de droite afin d'y reporter la ou les notes revues.

Exemple 1

Si vous avez attribué 10/20 à la première copie traitée, et que vous n'avez pas été amené à revoir la note de cette copie, il faut remplir le tableau de la manière suivante :

Rang de correction	Première note attribuée à la copie	Note sur 20 revue (le cas échéant)
1	10 / 20	

Exemple 2

Si vous avez attribué 8/20 à la vingtième copie traitée, mais qu'au cours ou à l'issue de la correction des copies suivantes, vous avez été amené à revoir la note de cette copie et à attribuer 11,5/20 puis finalement 12/20, il faut remplir le tableau de la manière suivante :

Rang de correction	Première note attribuée à la copie	Note sur 20 revue (le cas échéant)
20	8 / 20	11,5 puis 12

- FORMULAIRE DE COLLECTE DE DONNEES -

Date: / / 2015

Identifiant correcteur:

Correction en une seule fois : oui non

Années d'expérience:

Dans le tableau, veuillez reporter les notes attribuées au contrôle que vous êtes en train de corriger, **en respectant scrupuleusement l'ordre de correction des copies :**

1^{ère} copie corrigée = rang 1 ;

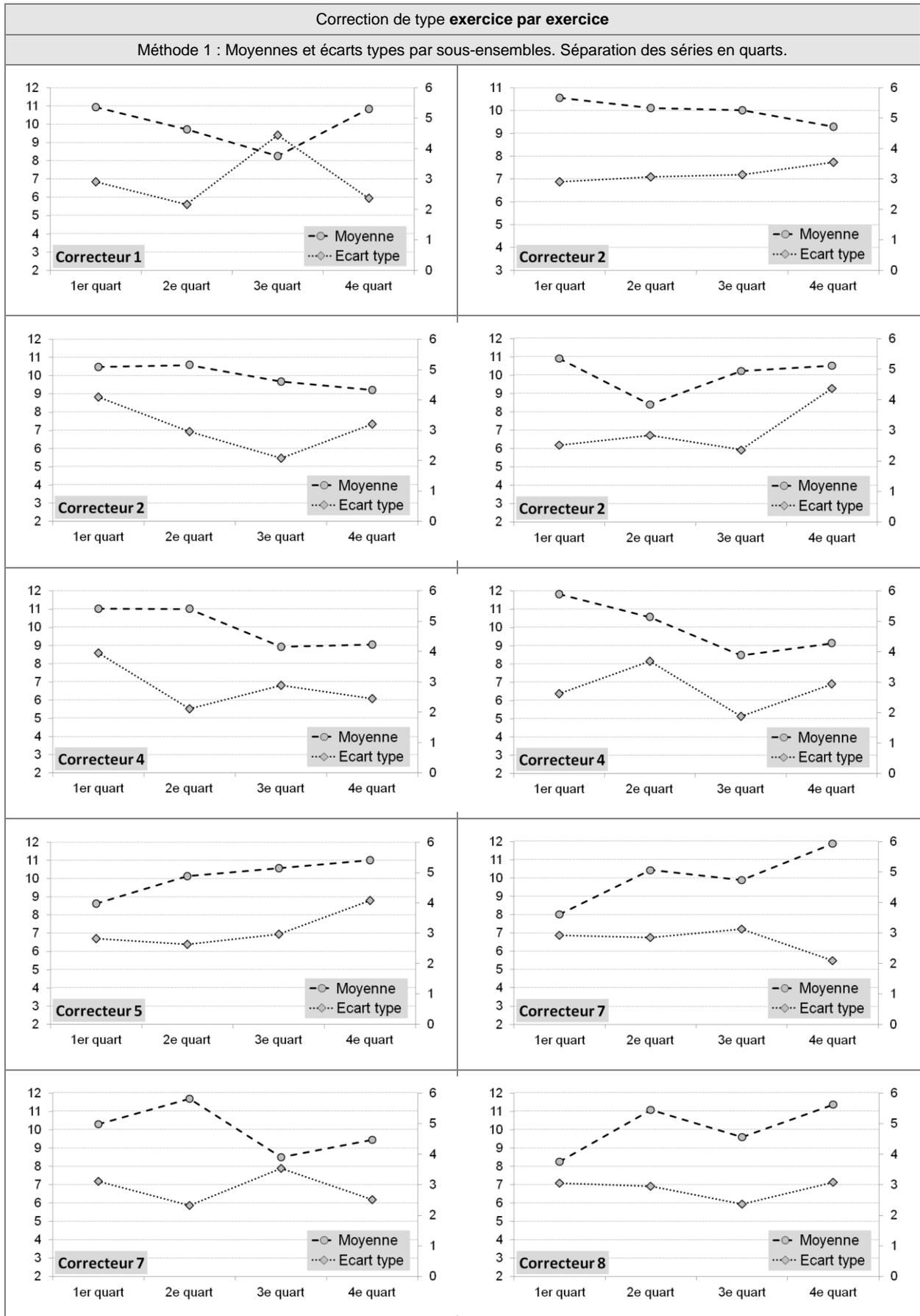
2nd copie corrigée = rang 2 ;

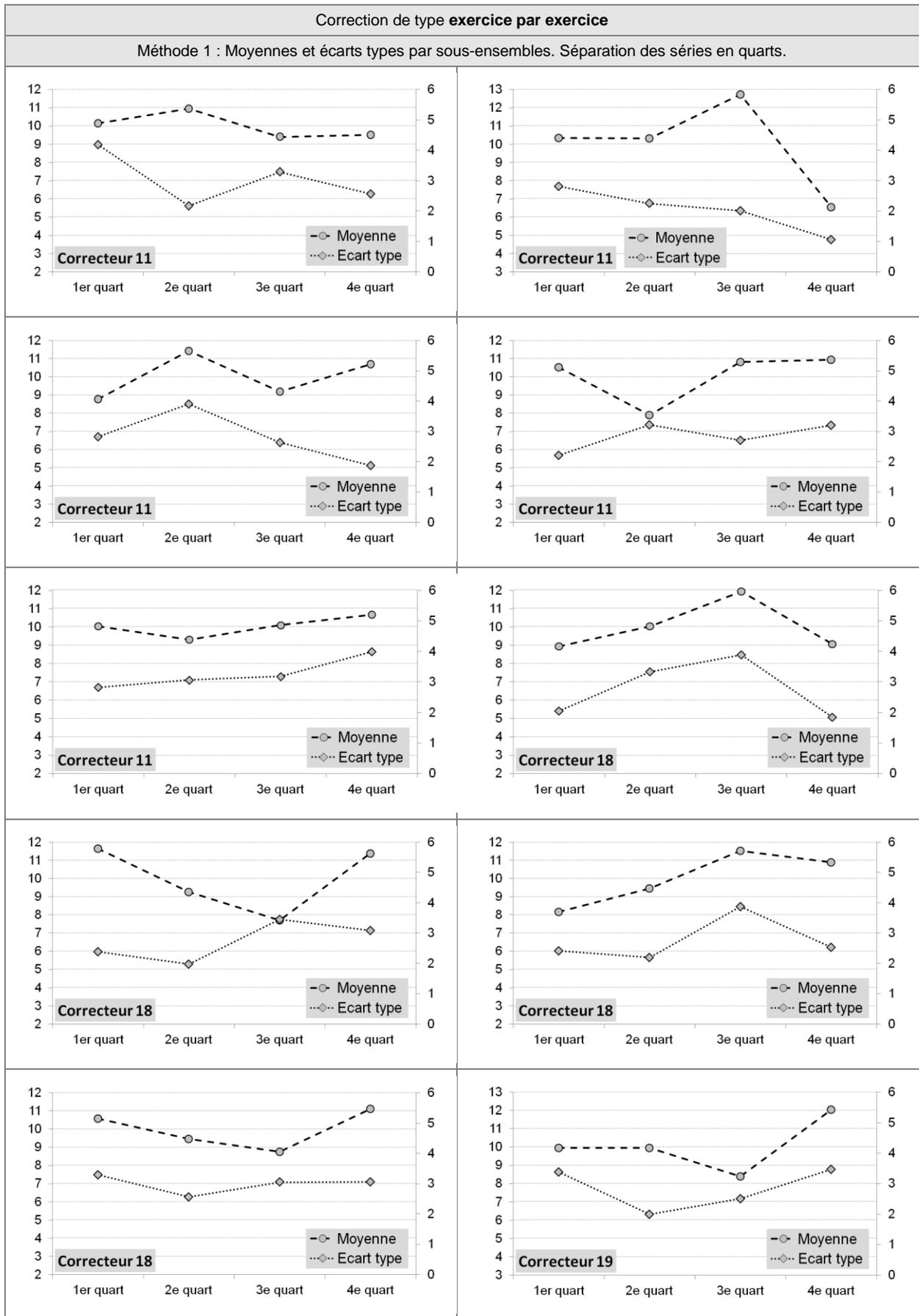
etc.

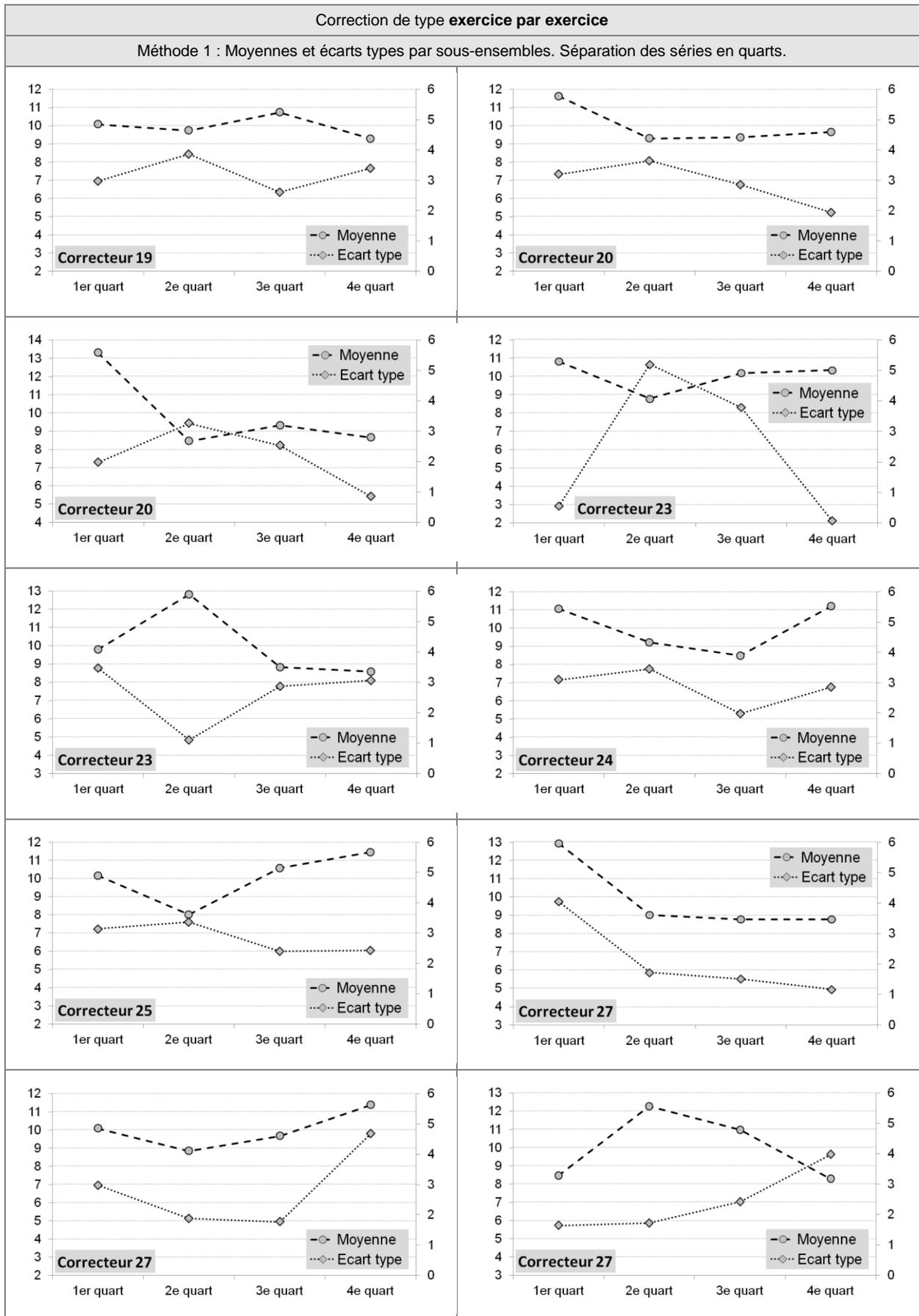
Ne triez surtout pas les notes dans l'ordre croissant ou décroissant.

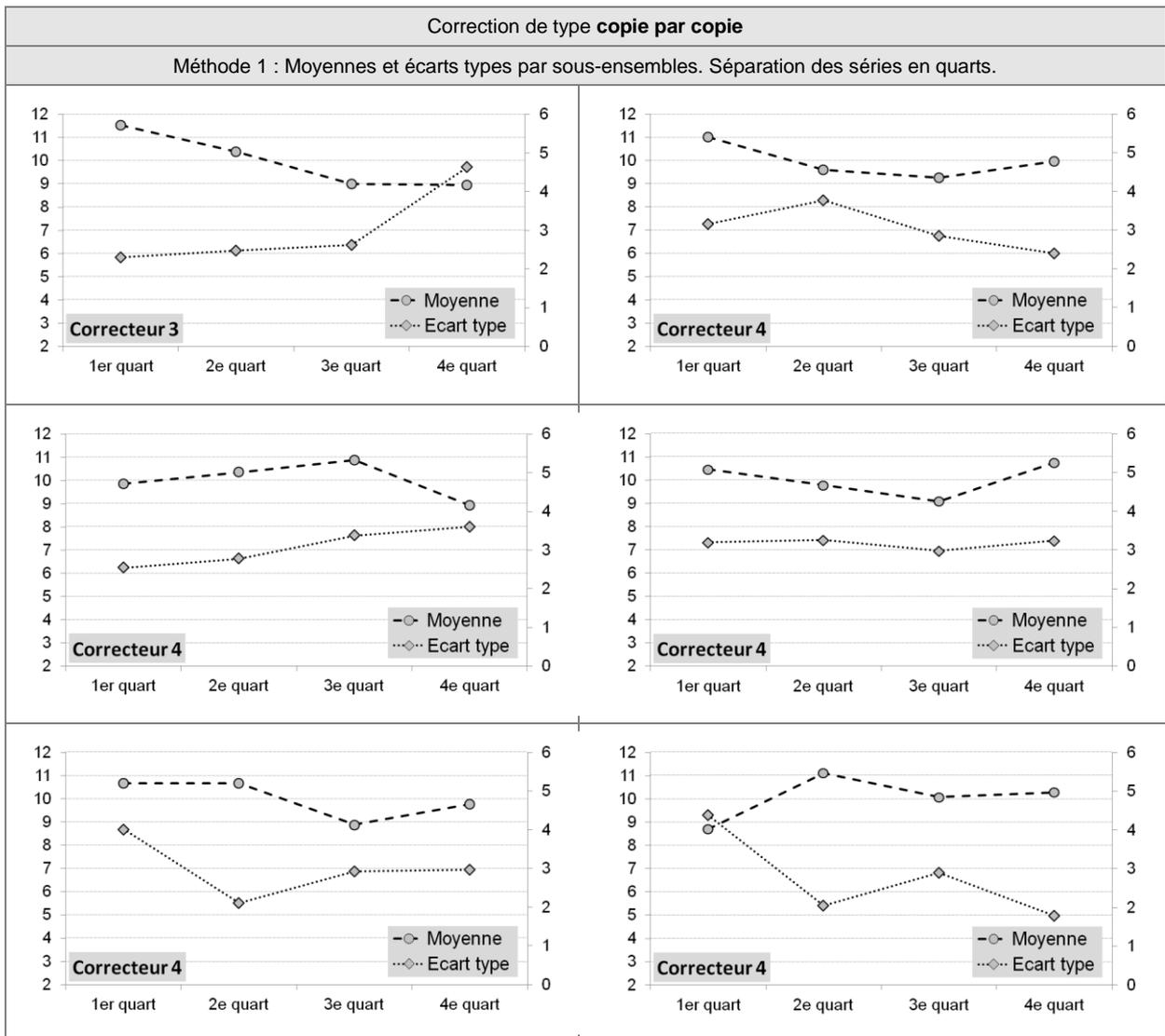
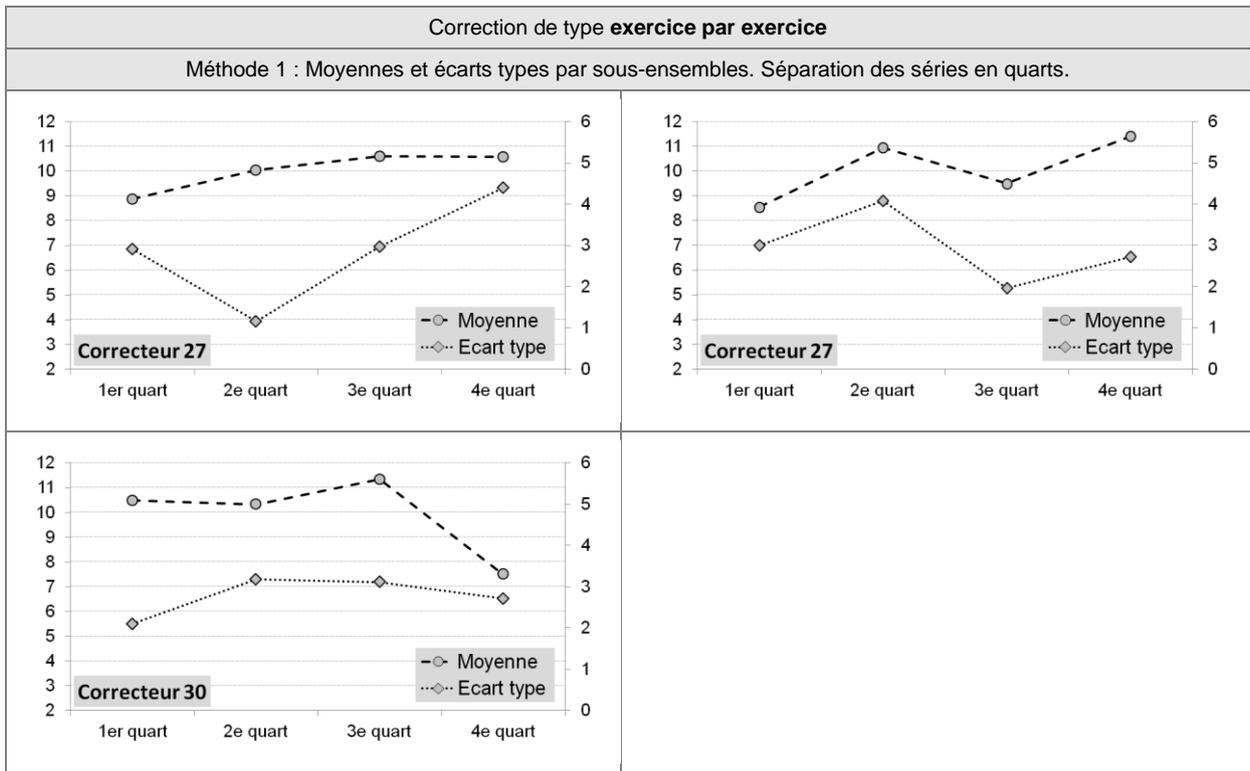
Rang de correction	Première note attribuée à la copie	Note sur 20 revue (le cas échéant)
1	/ 20	
2	/ 20	
3	/ 20	
4	/ 20	
5	/ 20	
6	/ 20	
7	/ 20	
8	/ 20	
9	/ 20	
10	/ 20	
11	/ 20	
12	/ 20	
13	/ 20	
14	/ 20	
15	/ 20	
16	/ 20	
17	/ 20	
18	/ 20	
19	/ 20	
20	/ 20	
21	/ 20	
22	/ 20	
23	/ 20	
24	/ 20	
25	/ 20	
26	/ 20	
27	/ 20	
28	/ 20	
29	/ 20	
30	/ 20	
31	/ 20	
32	/ 20	
33	/ 20	
34	/ 20	
35	/ 20	
36	/ 20	
37	/ 20	

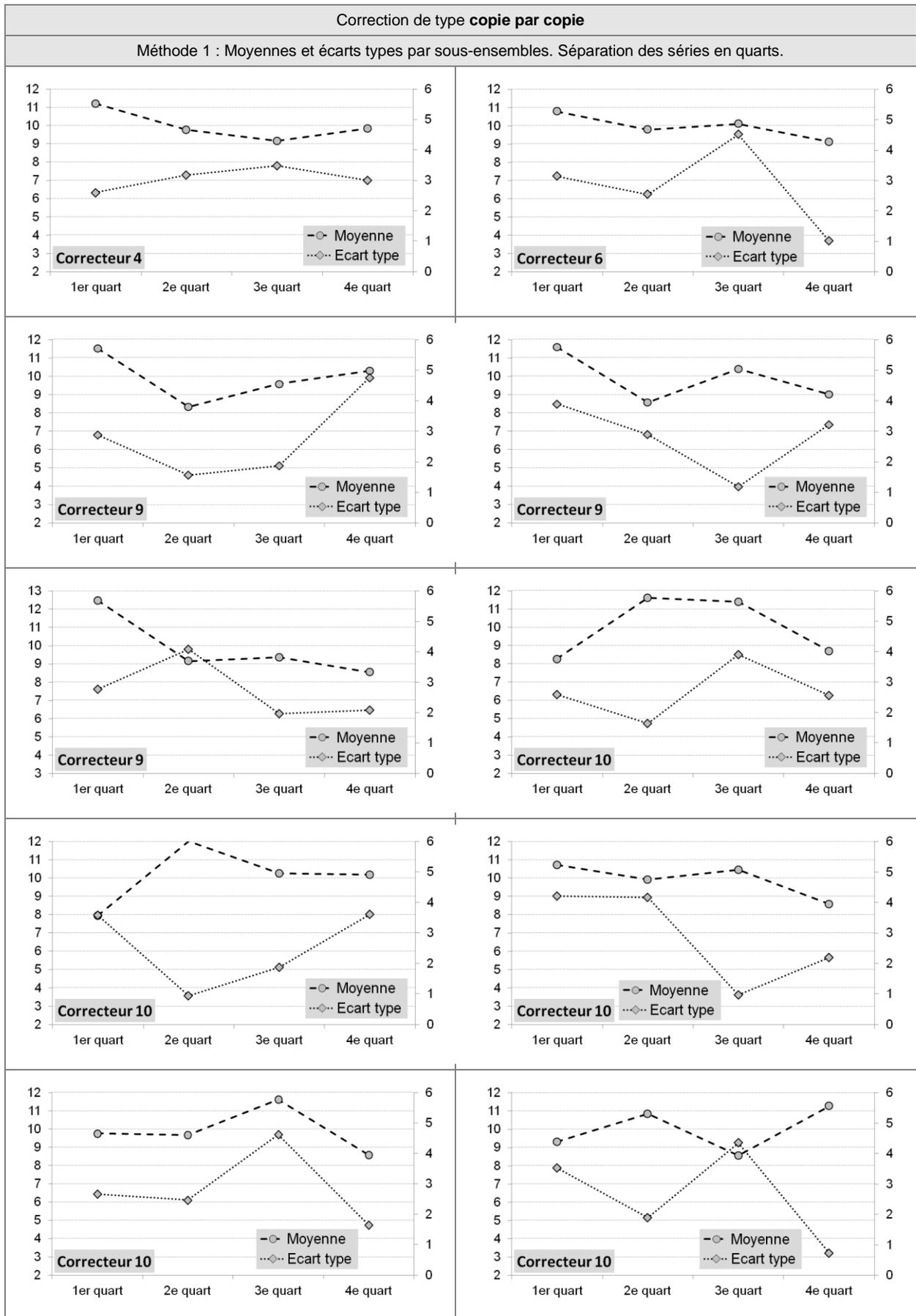
Annexe 2: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en quarts.

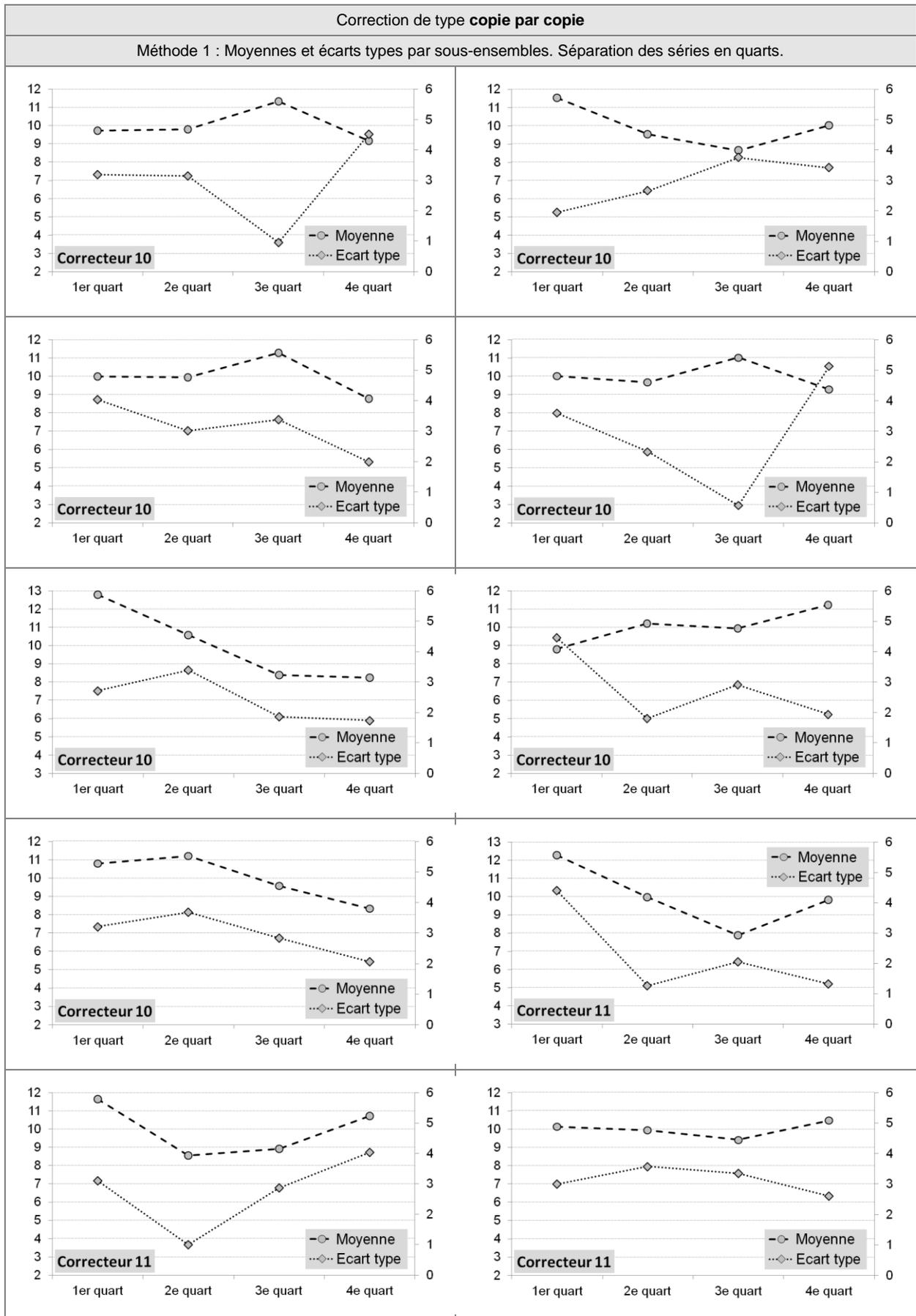


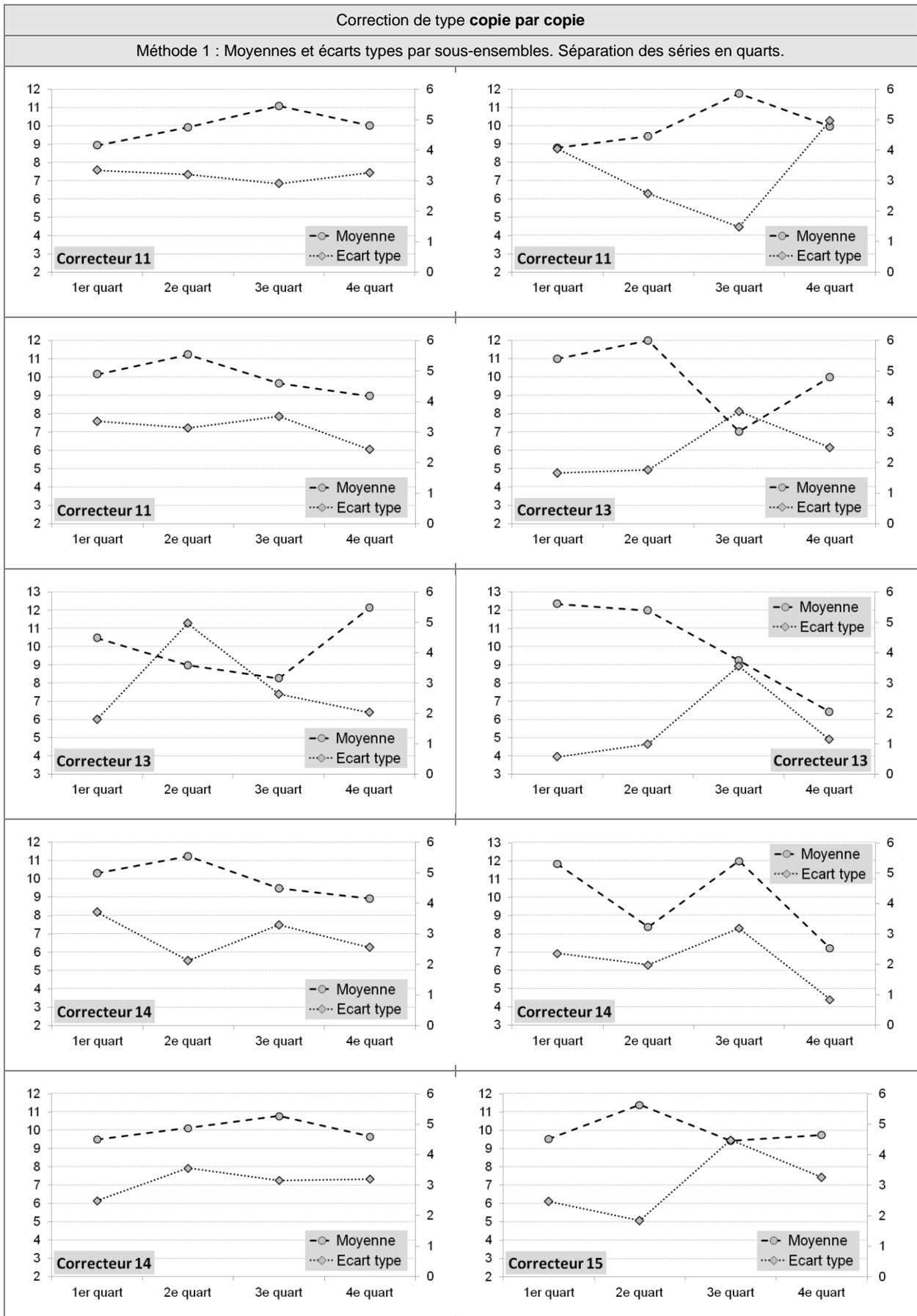


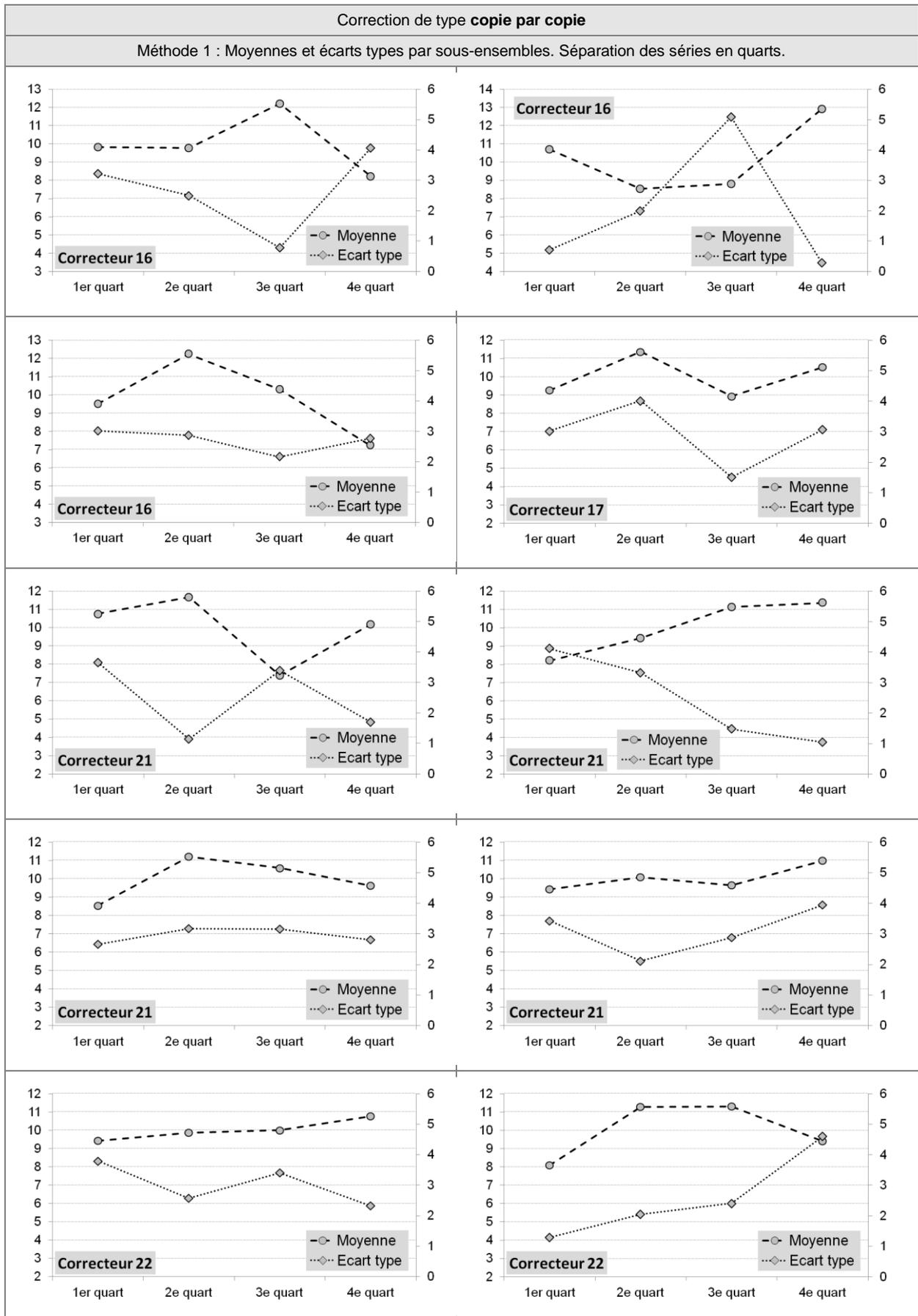


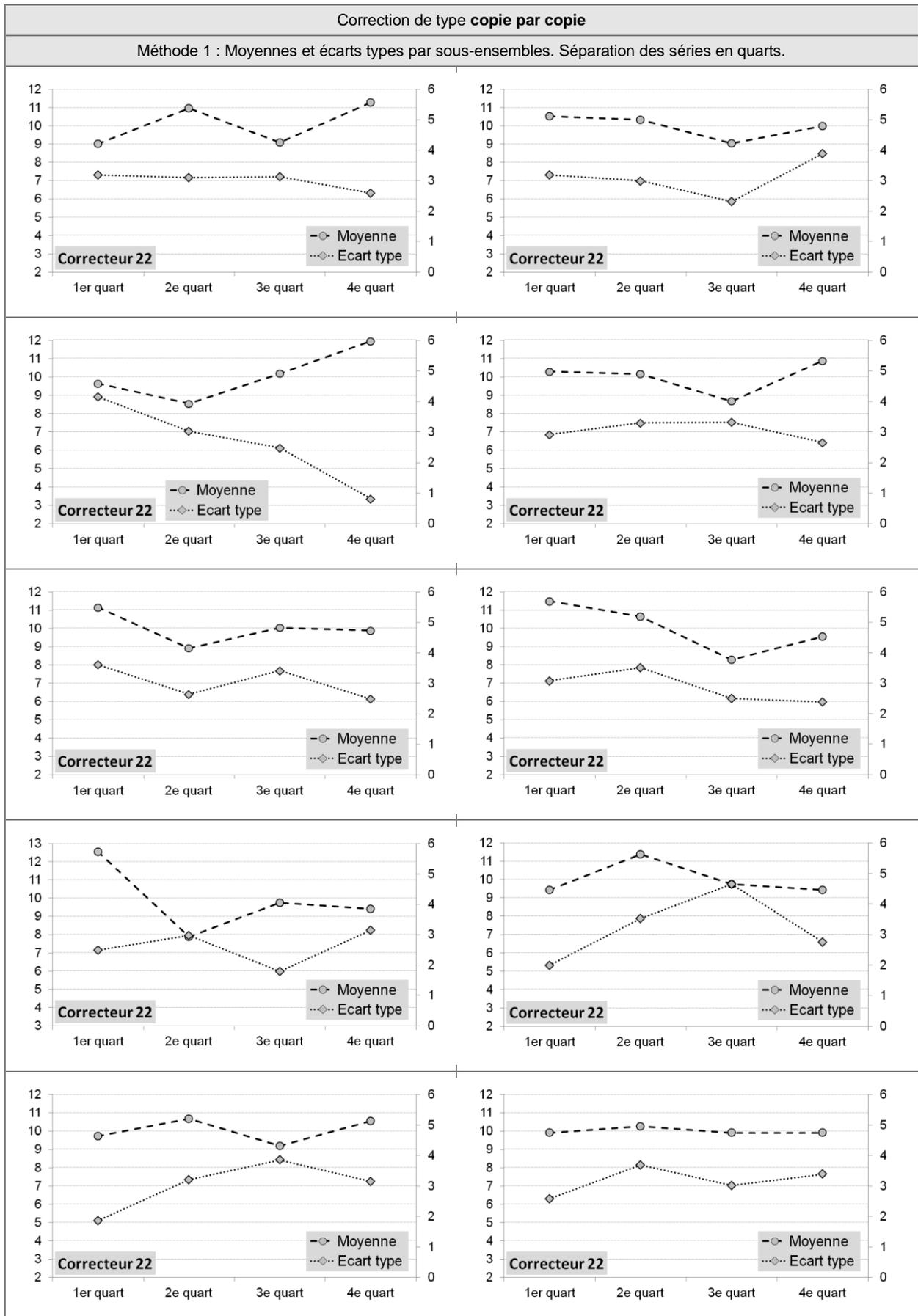


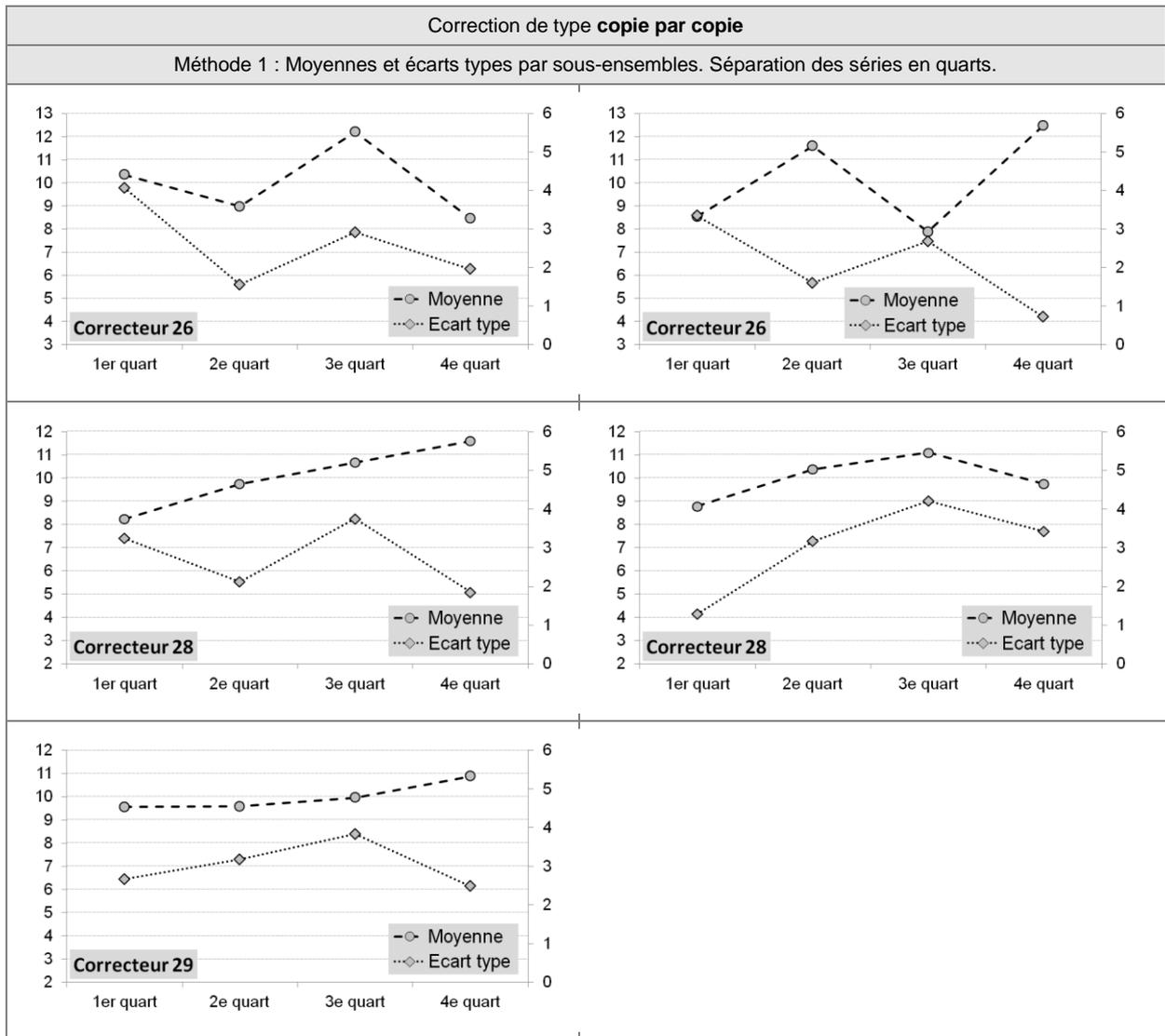




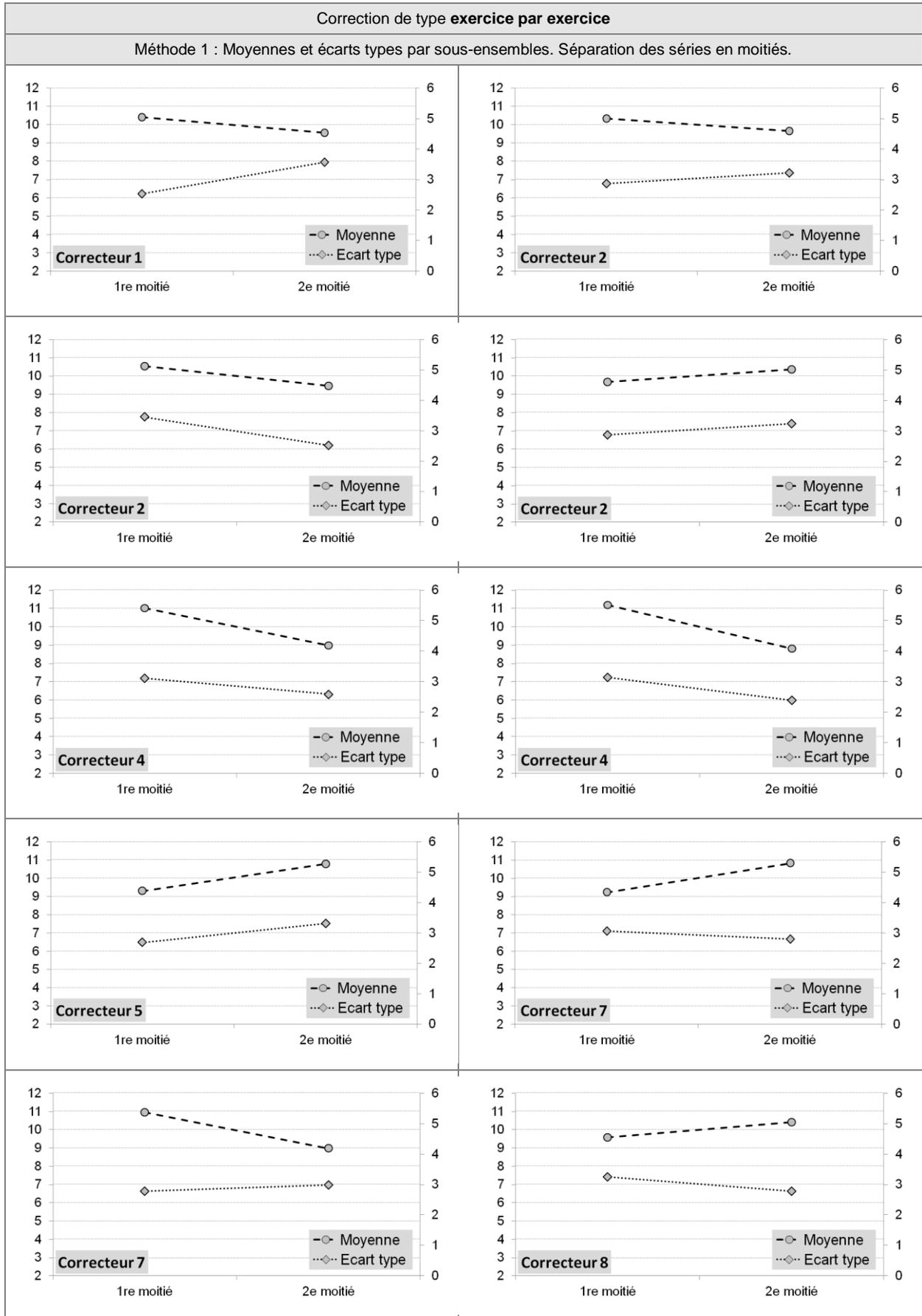


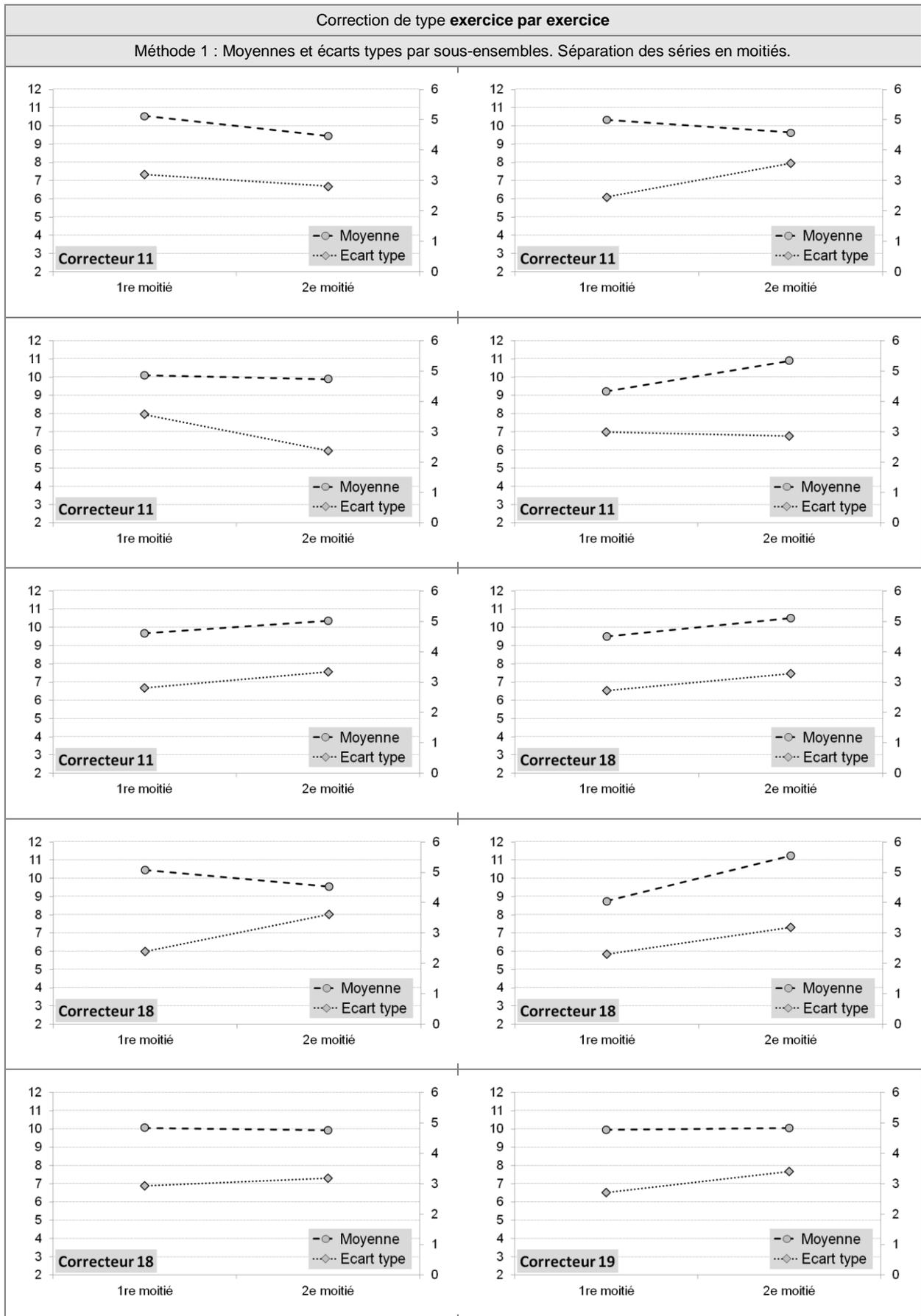


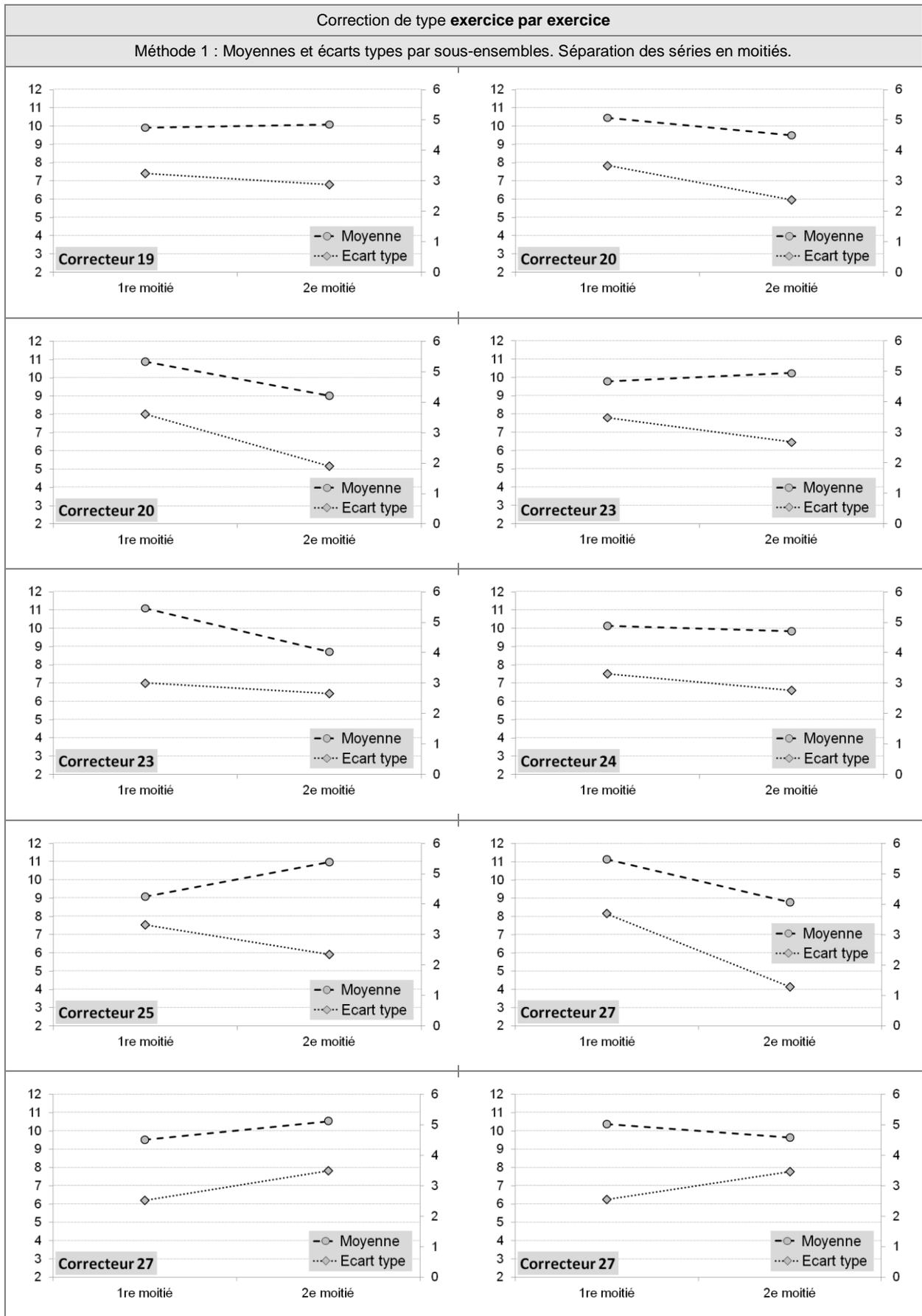


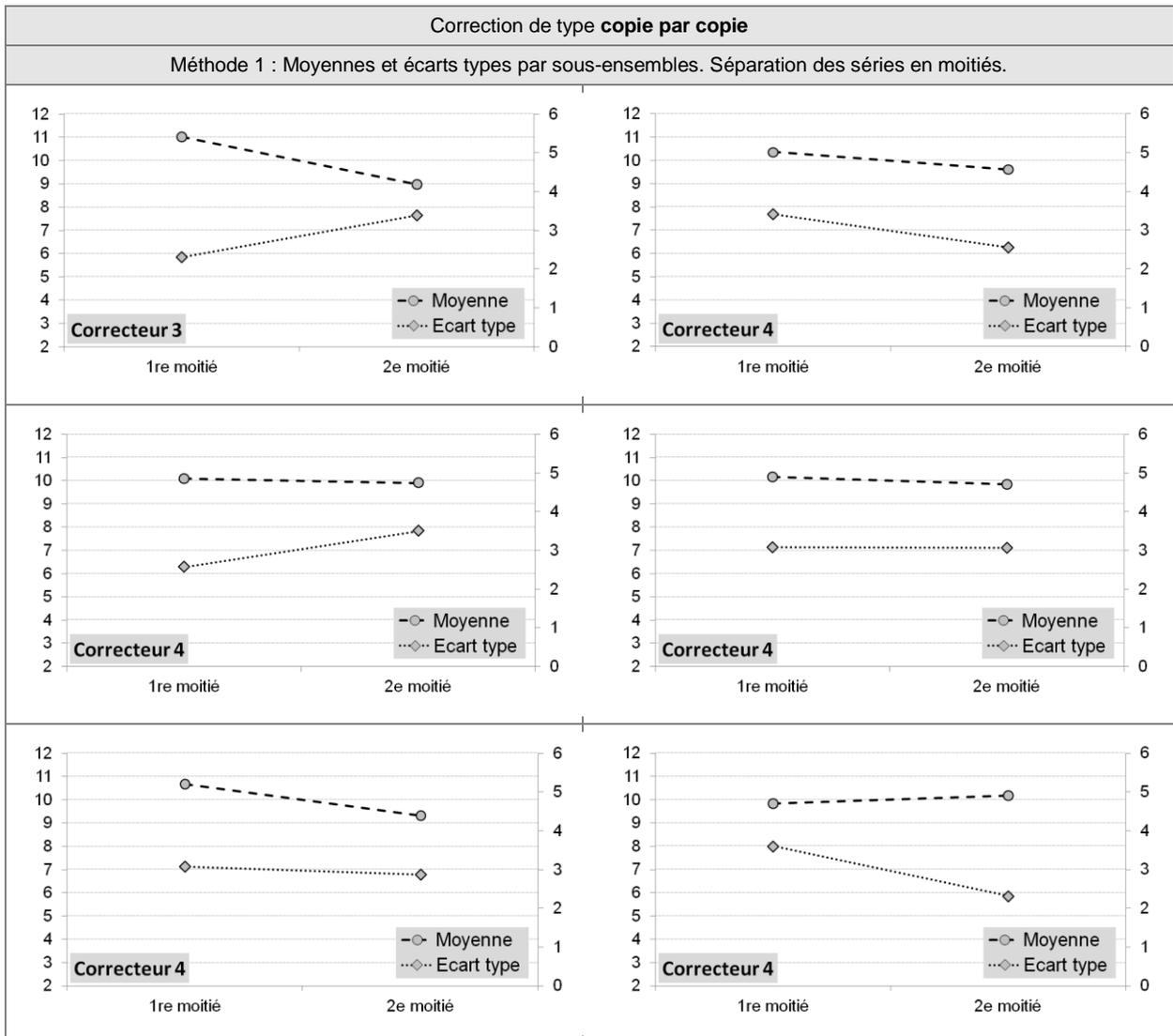
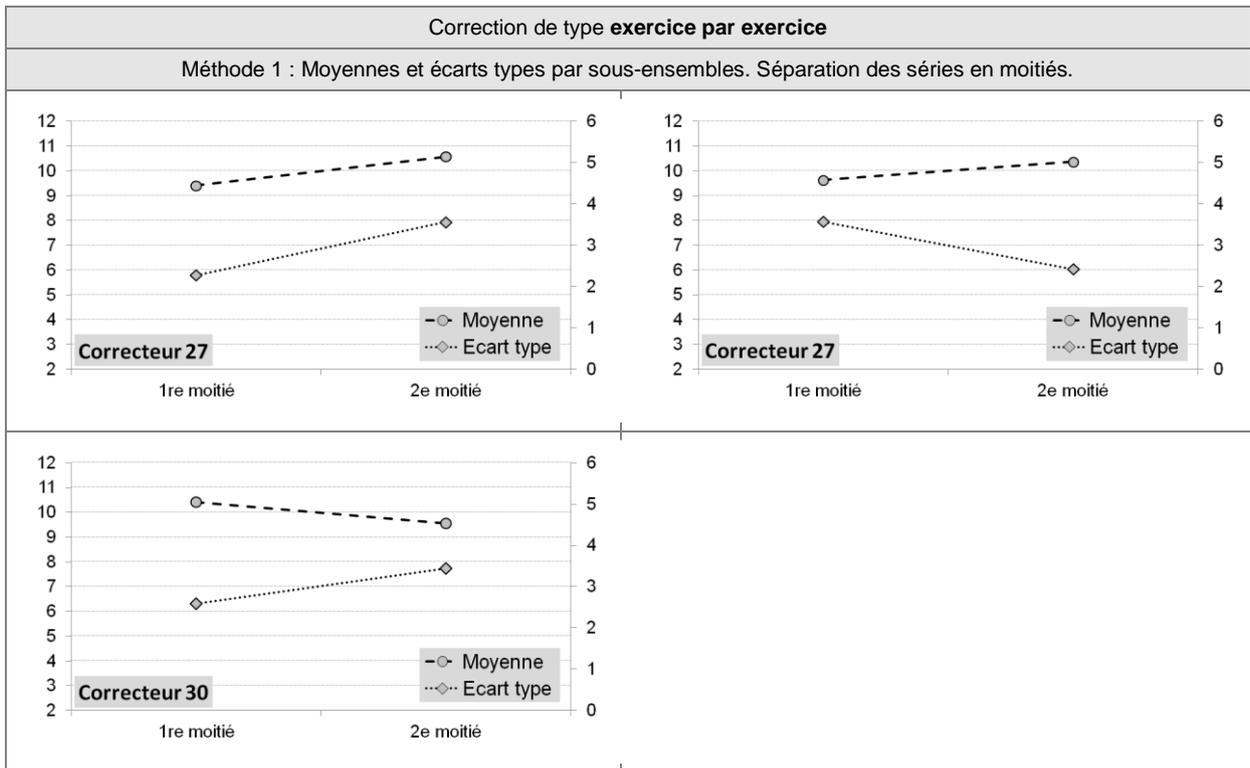


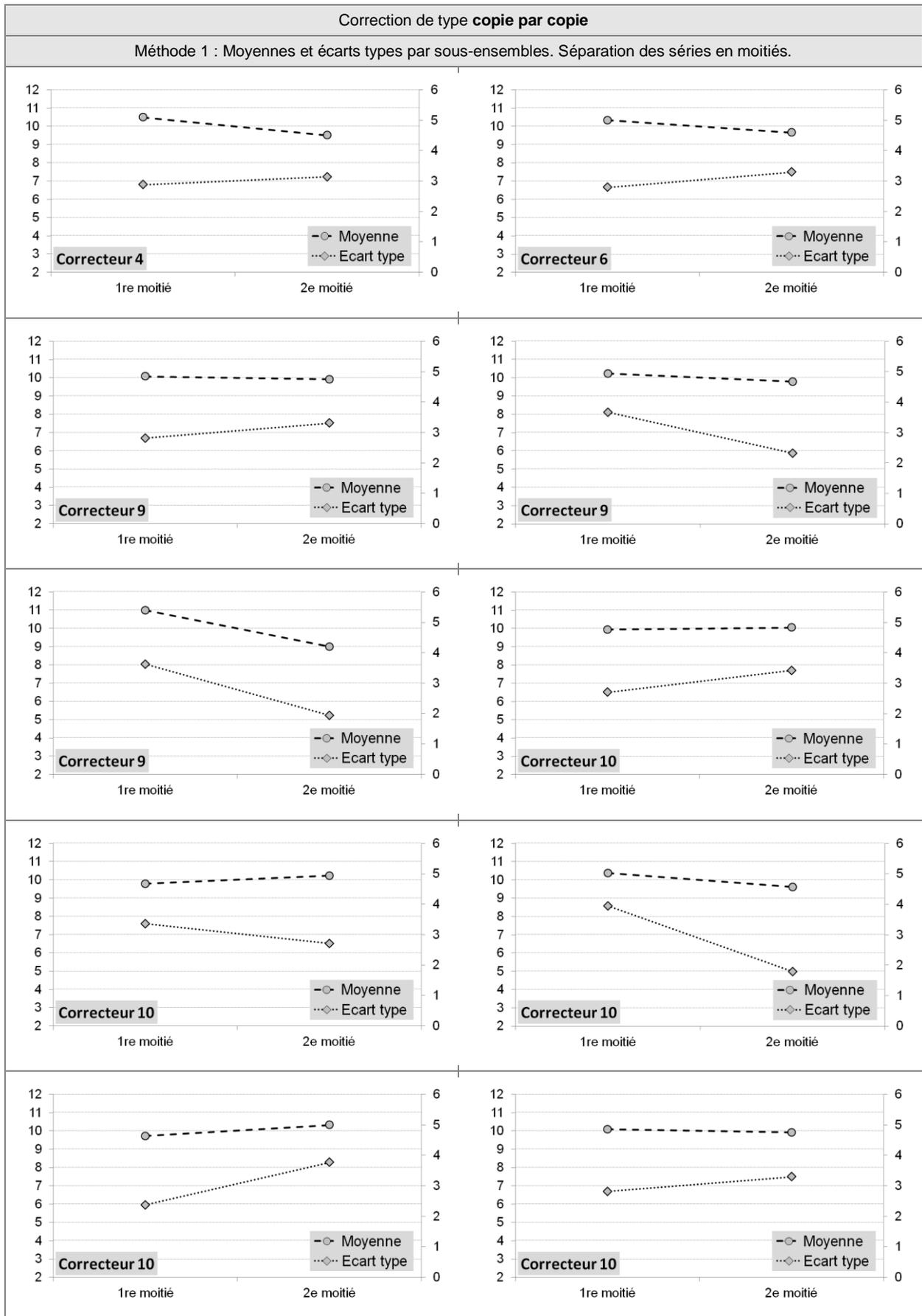
Annexe 3: Planches de résultats obtenus par la méthode 1 : moyennes et écarts types par sous-ensembles. Séparation des séries en moitiés.

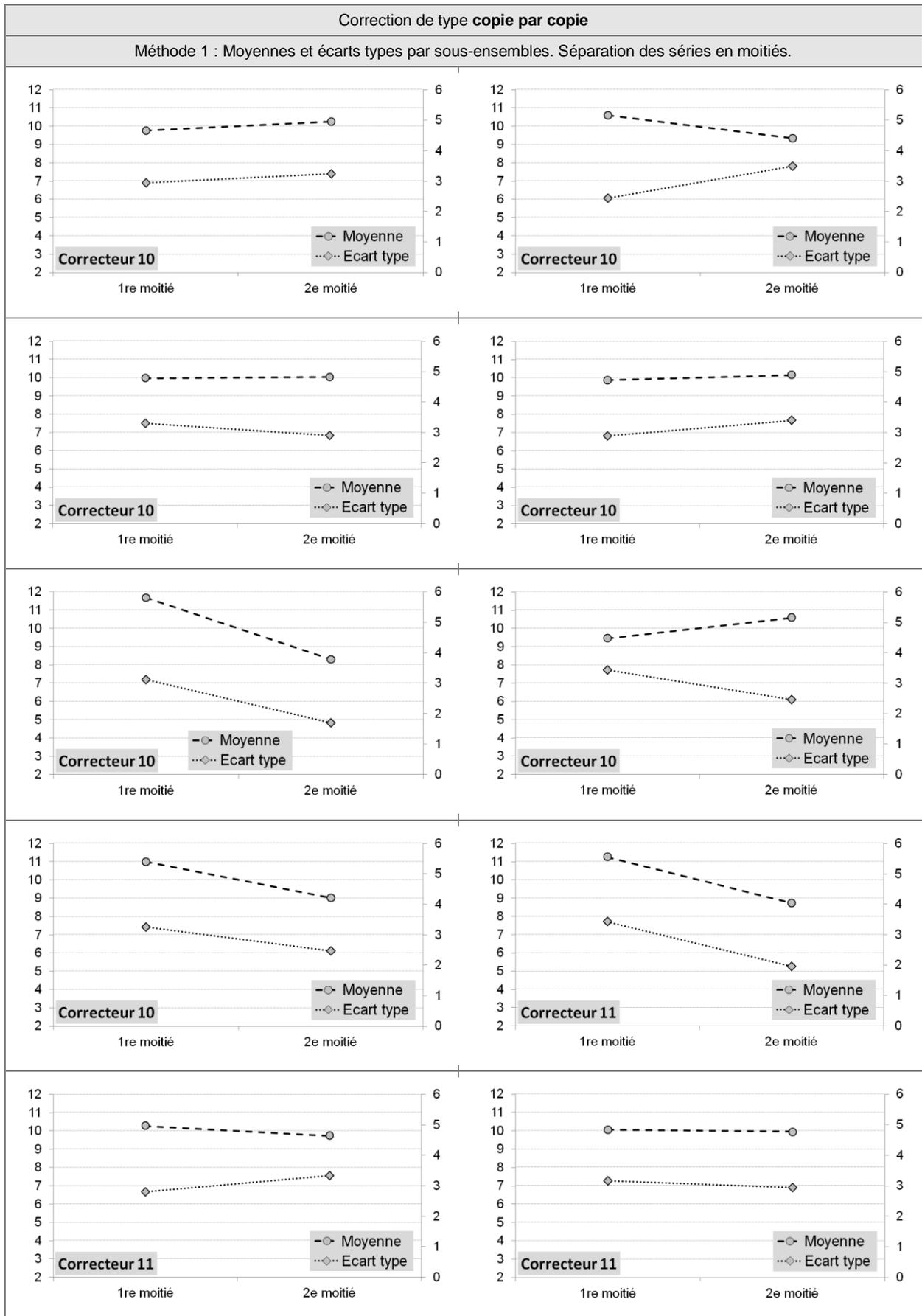


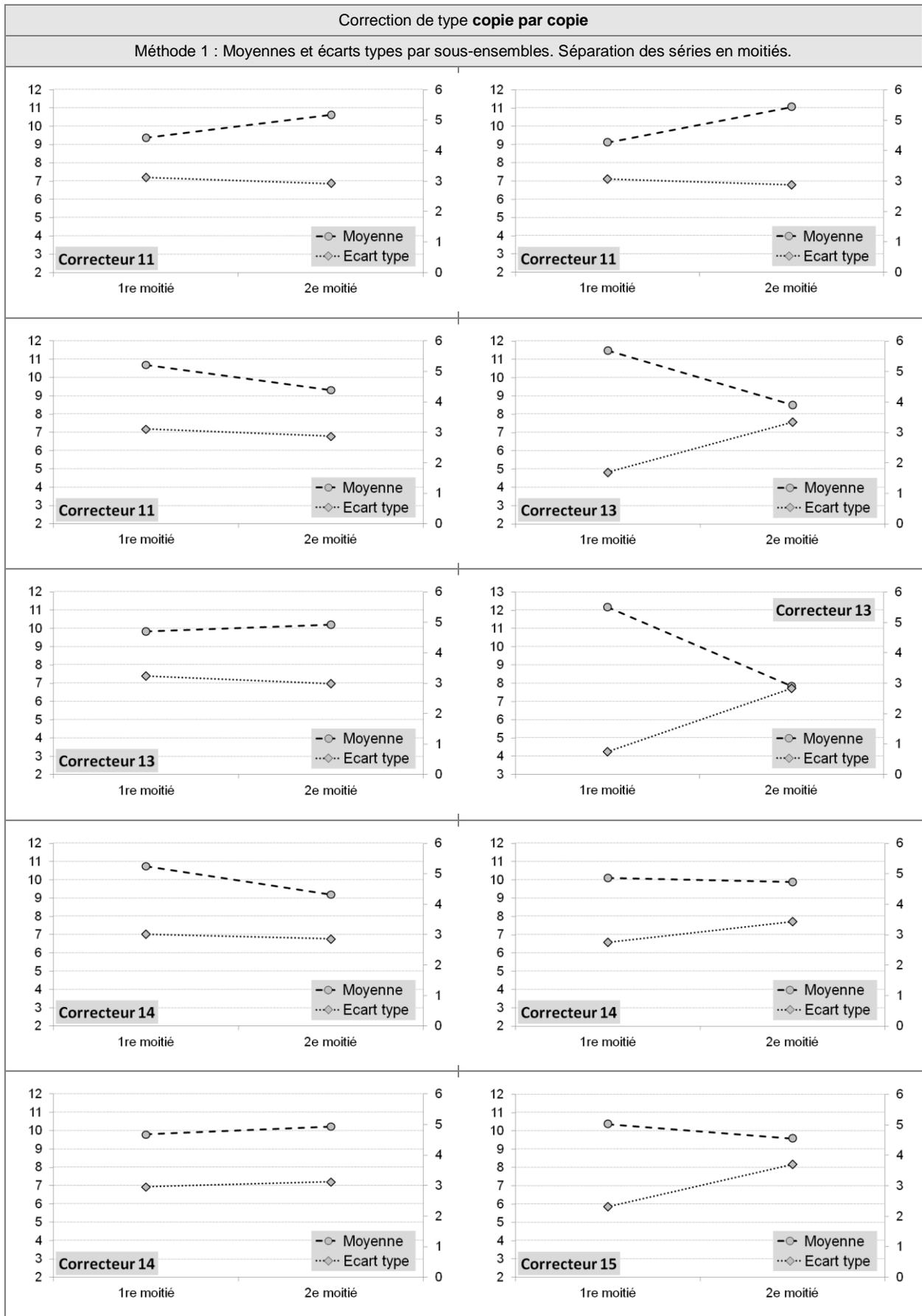


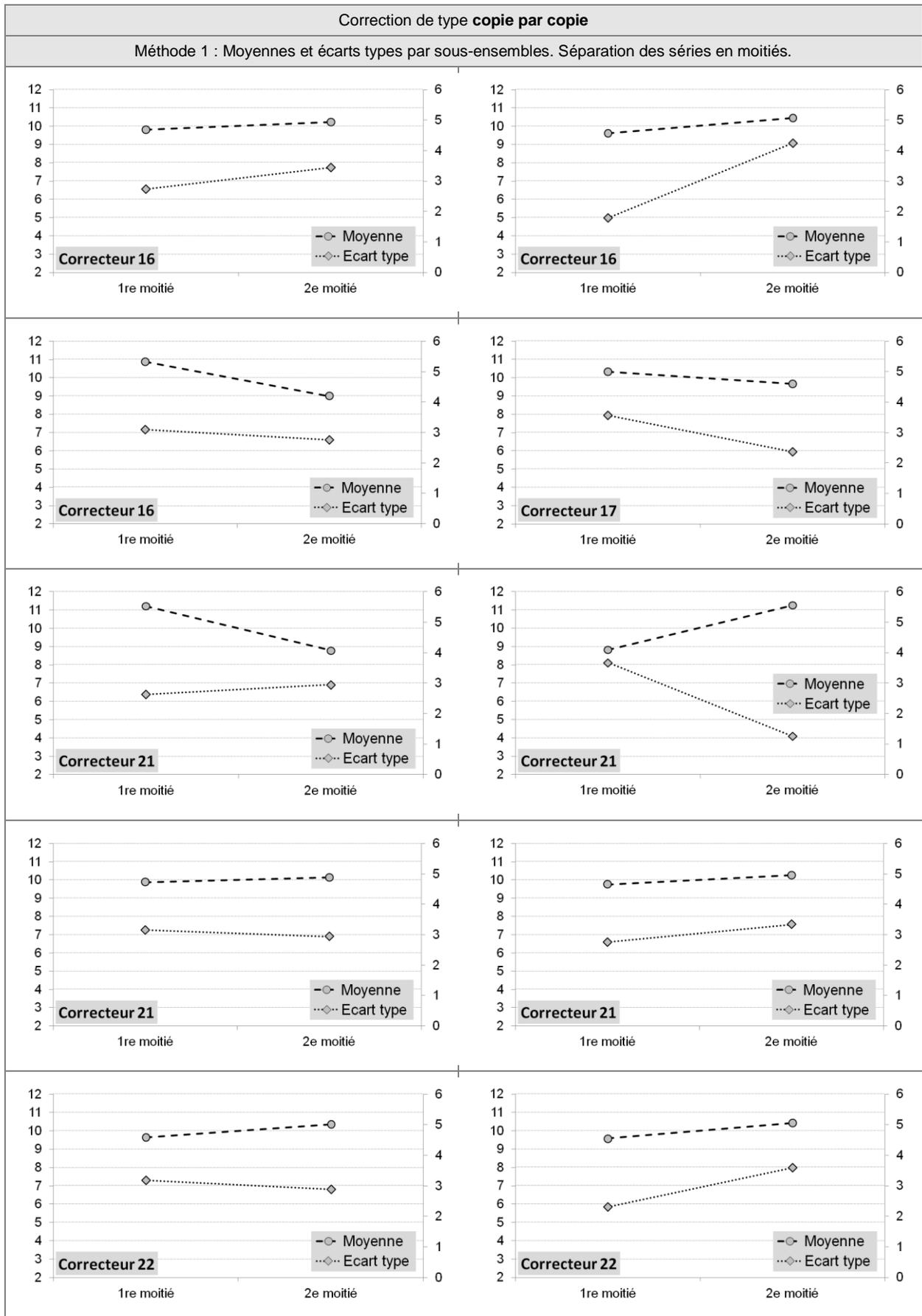


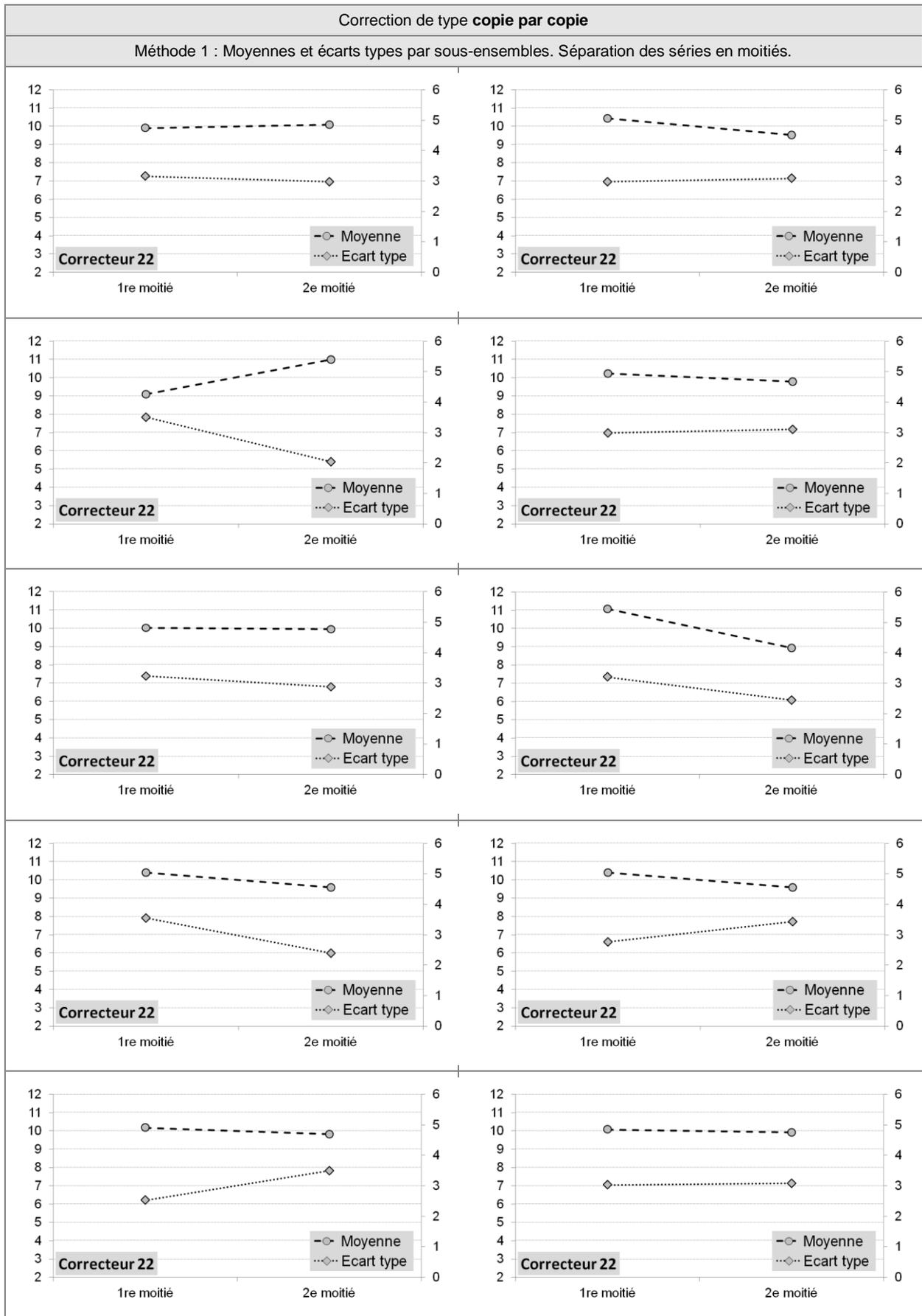


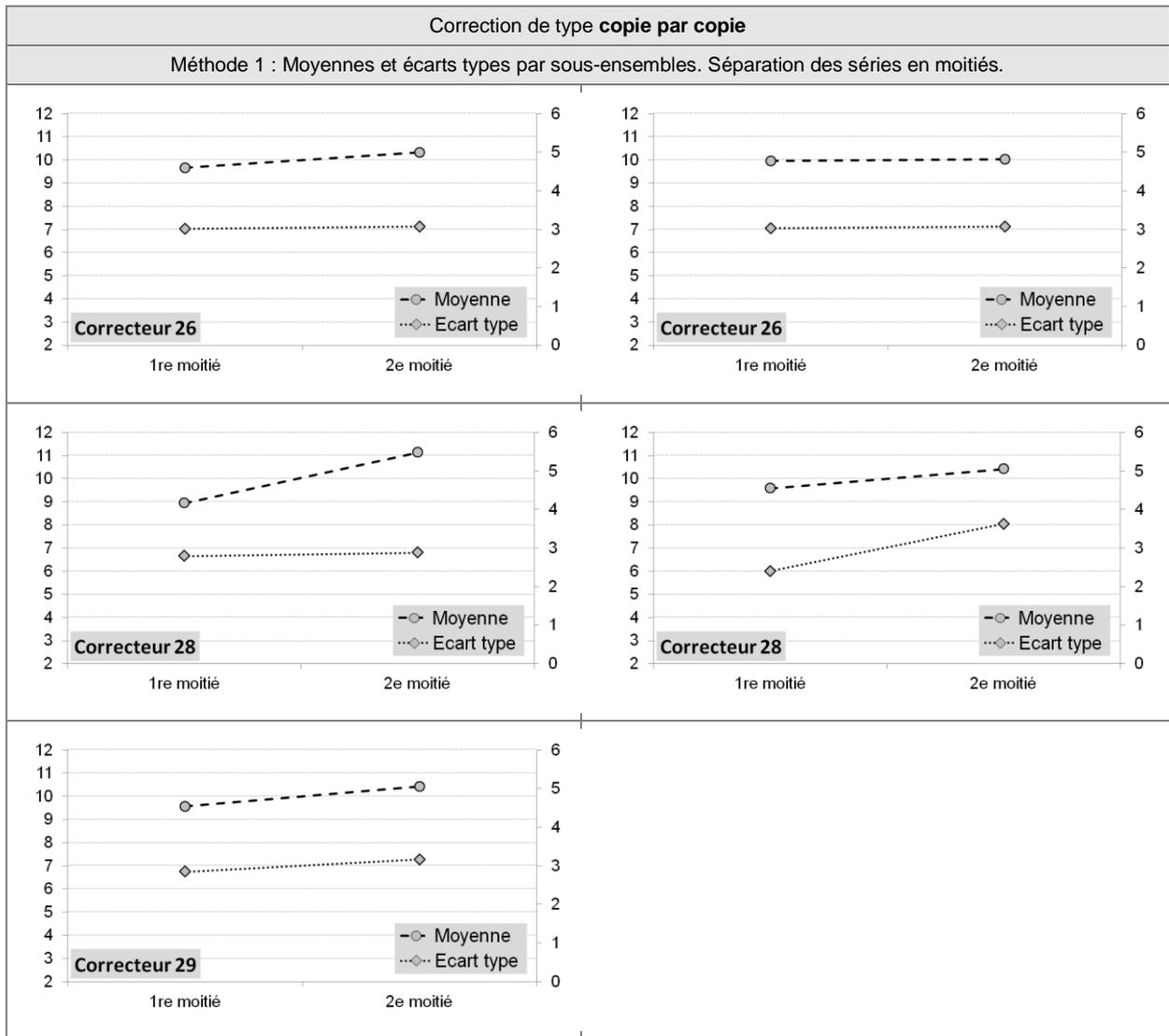




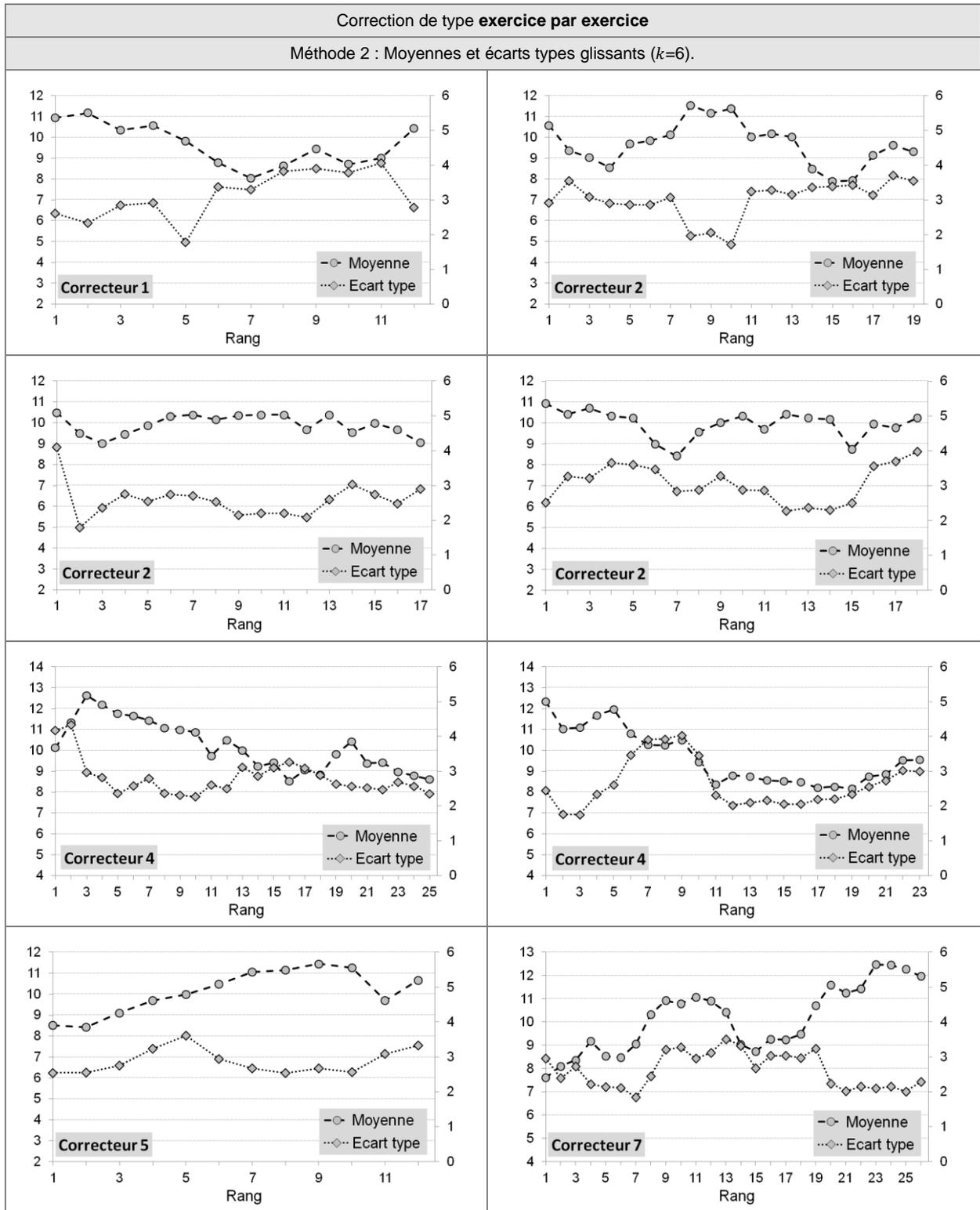


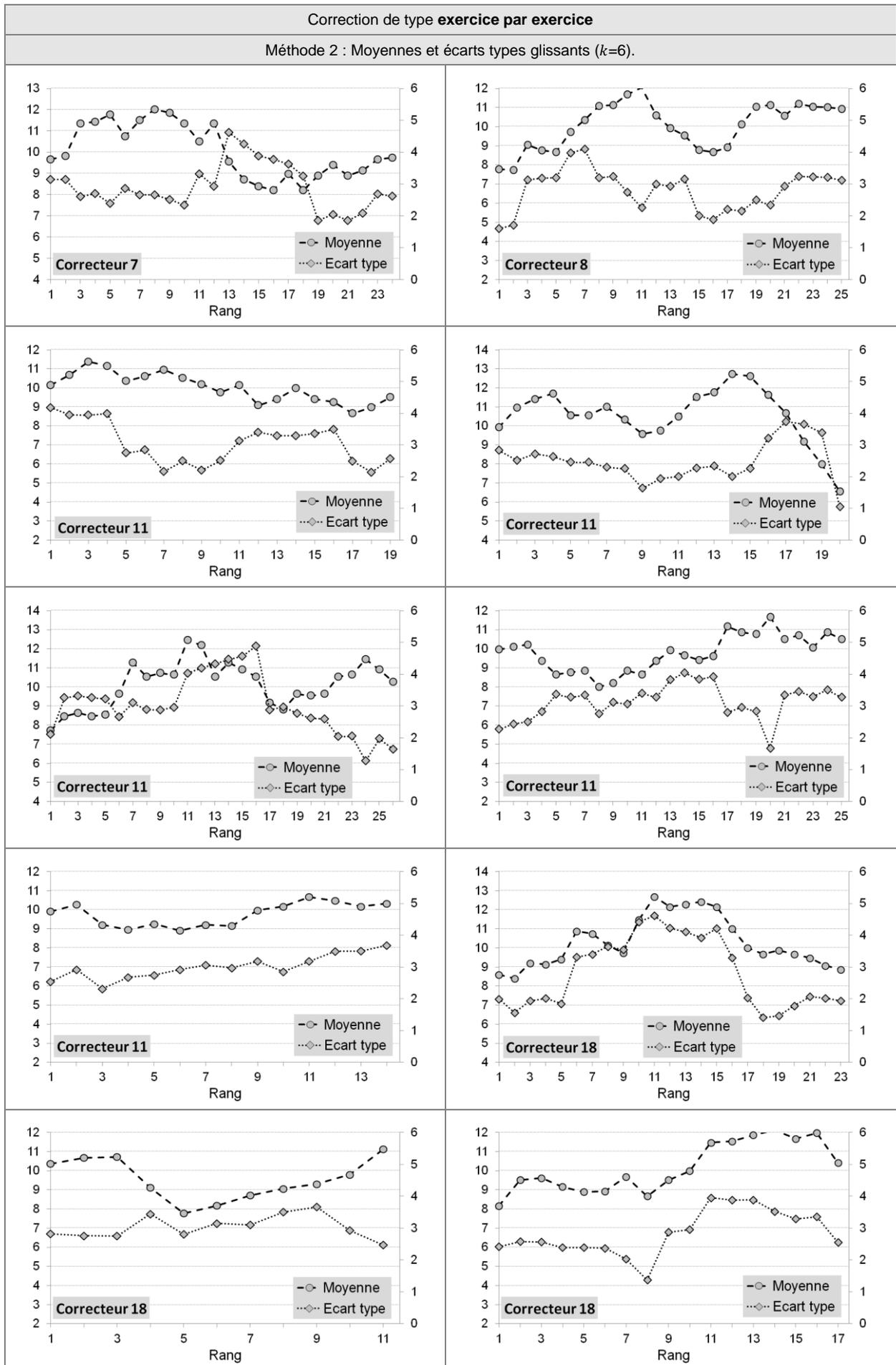


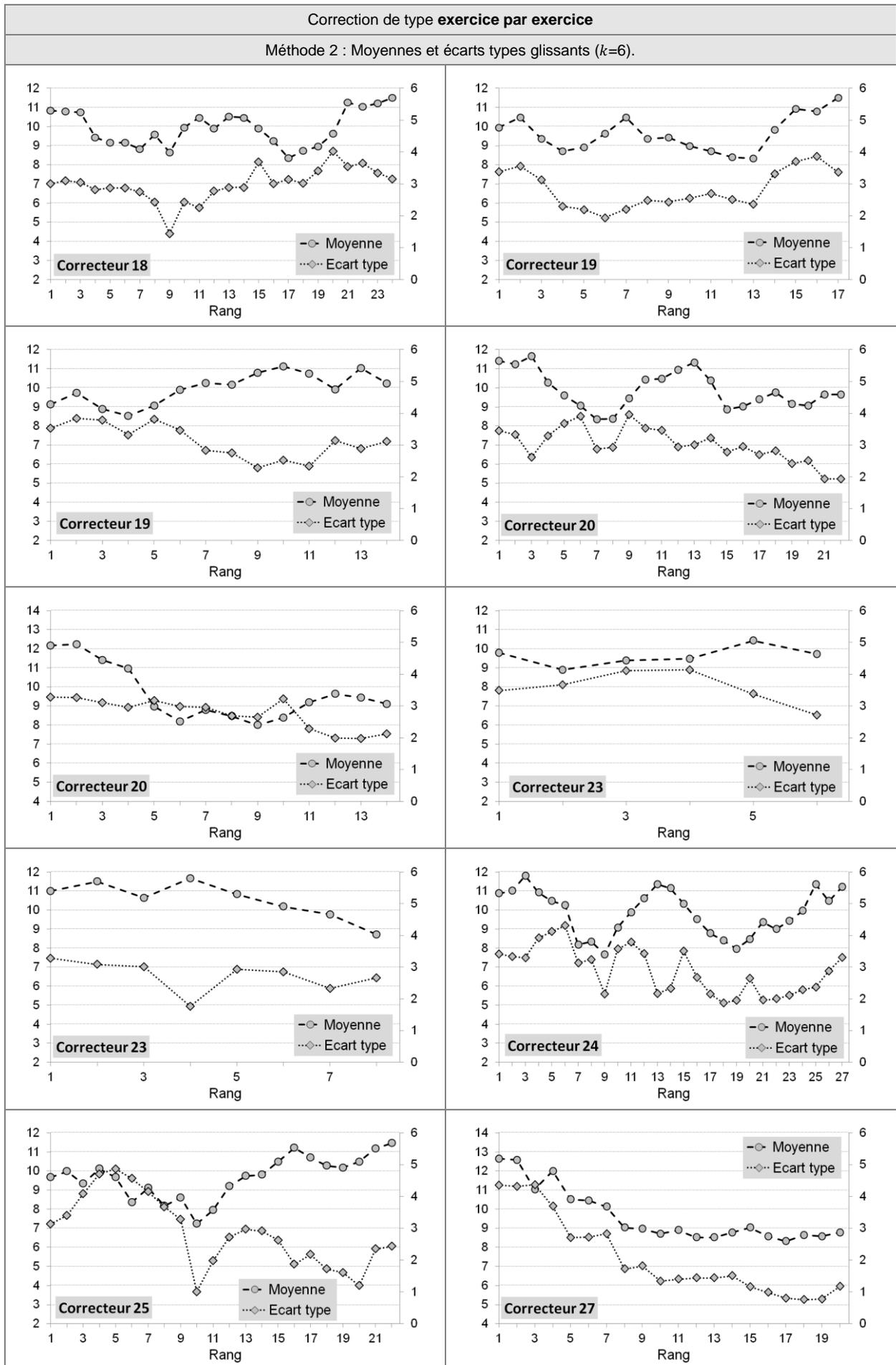


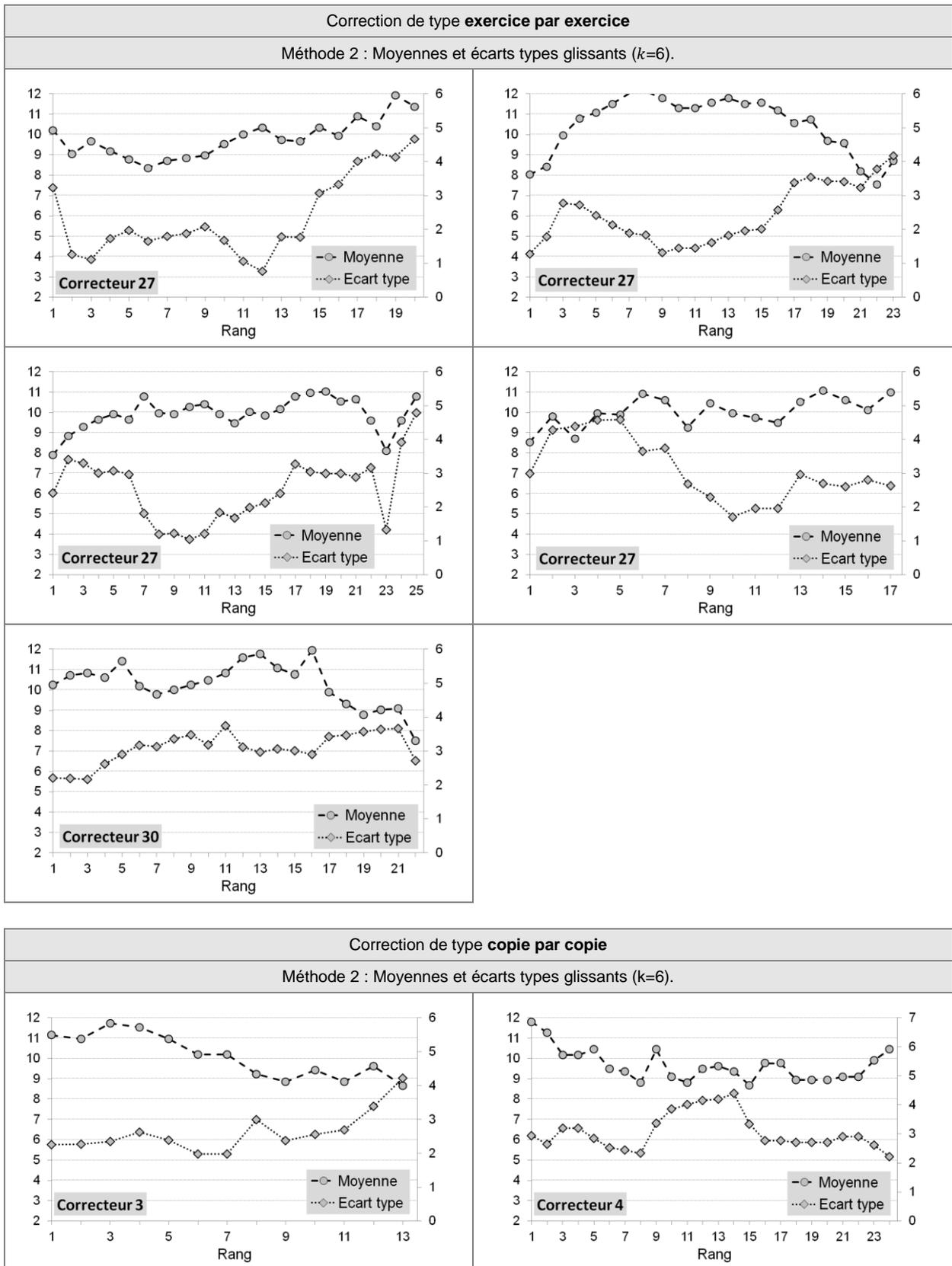


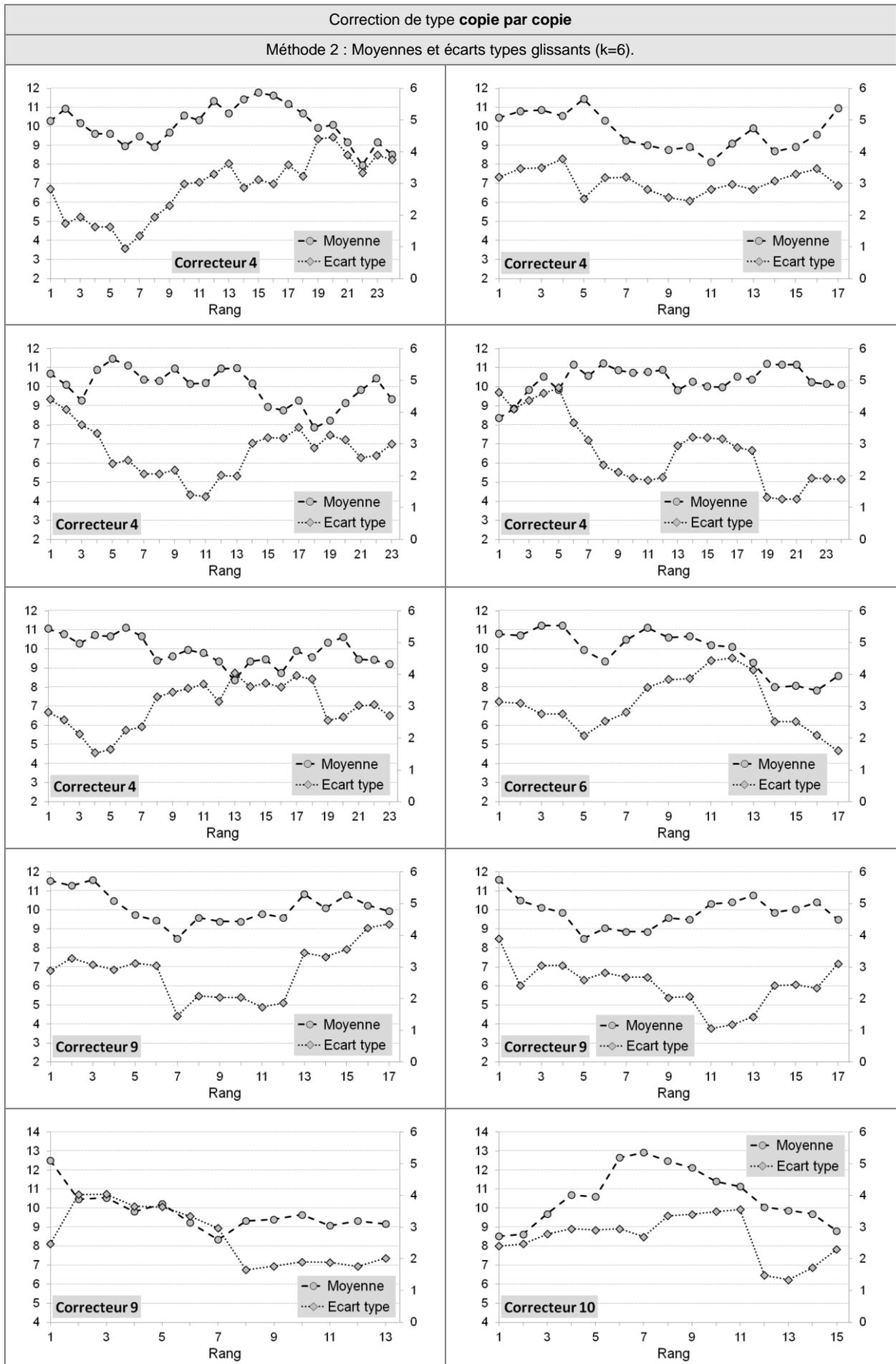
Annexe 4: Planches de résultats obtenus par la méthode 2 : moyennes et écarts types glissants.

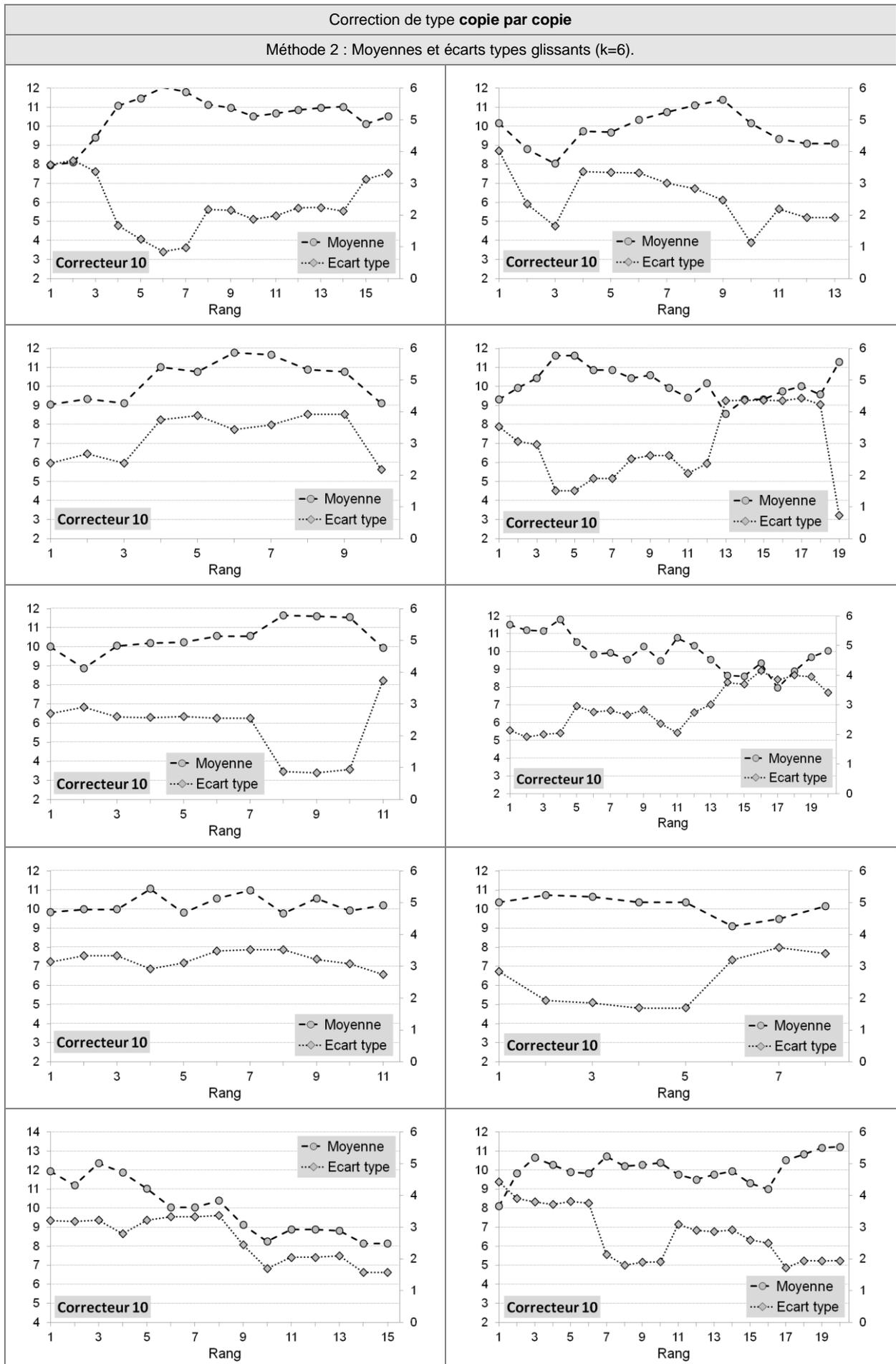


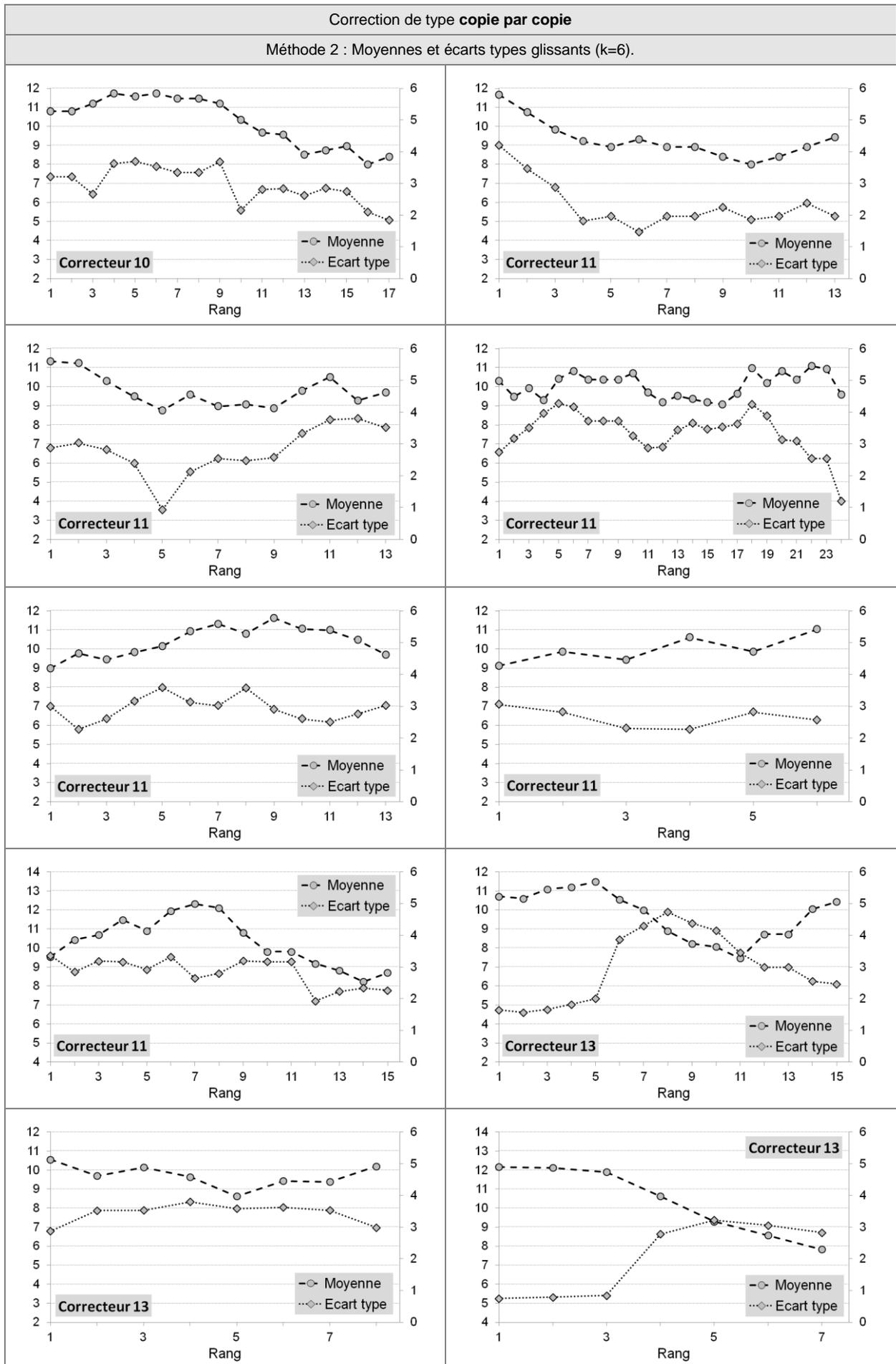


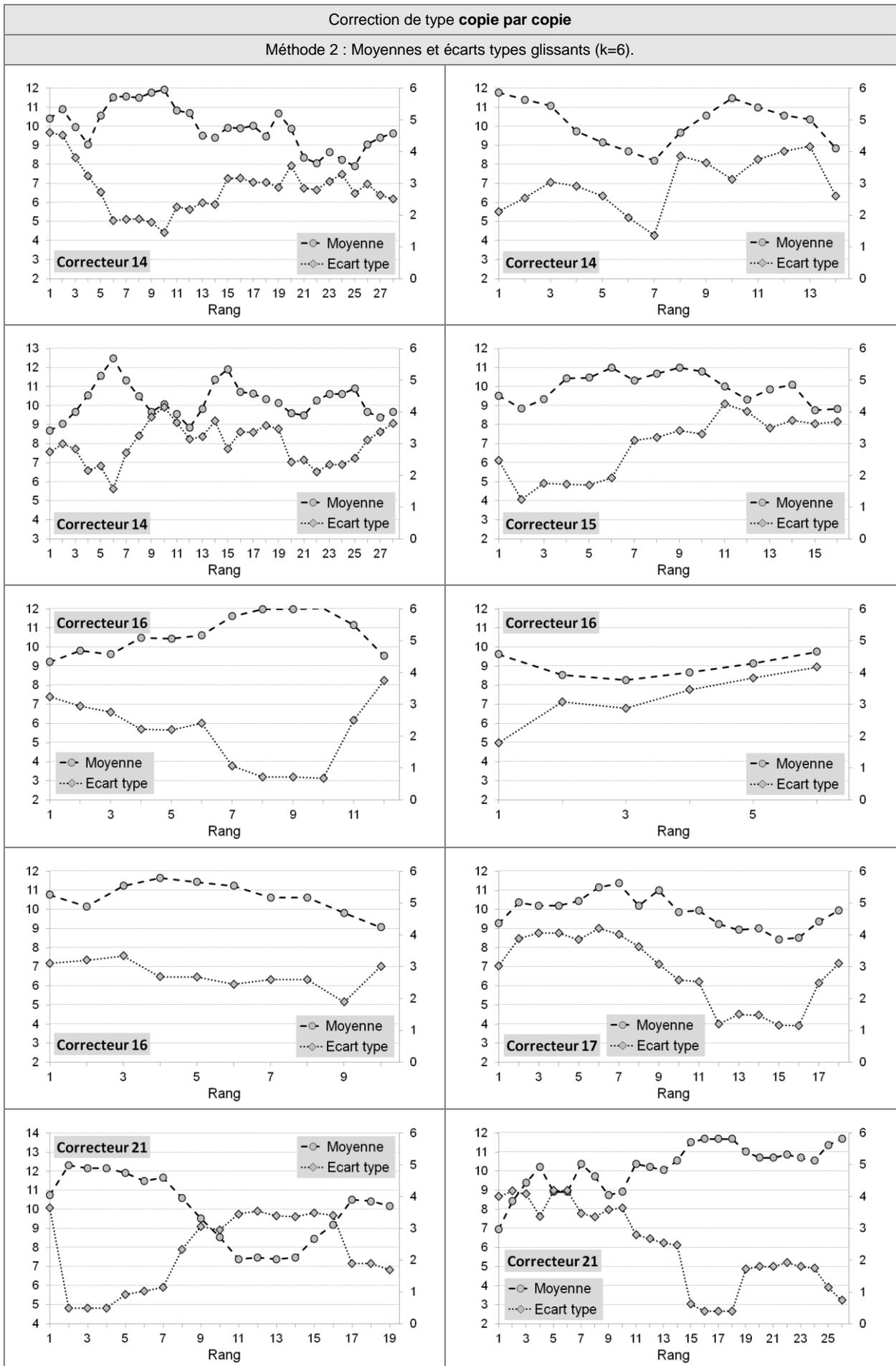


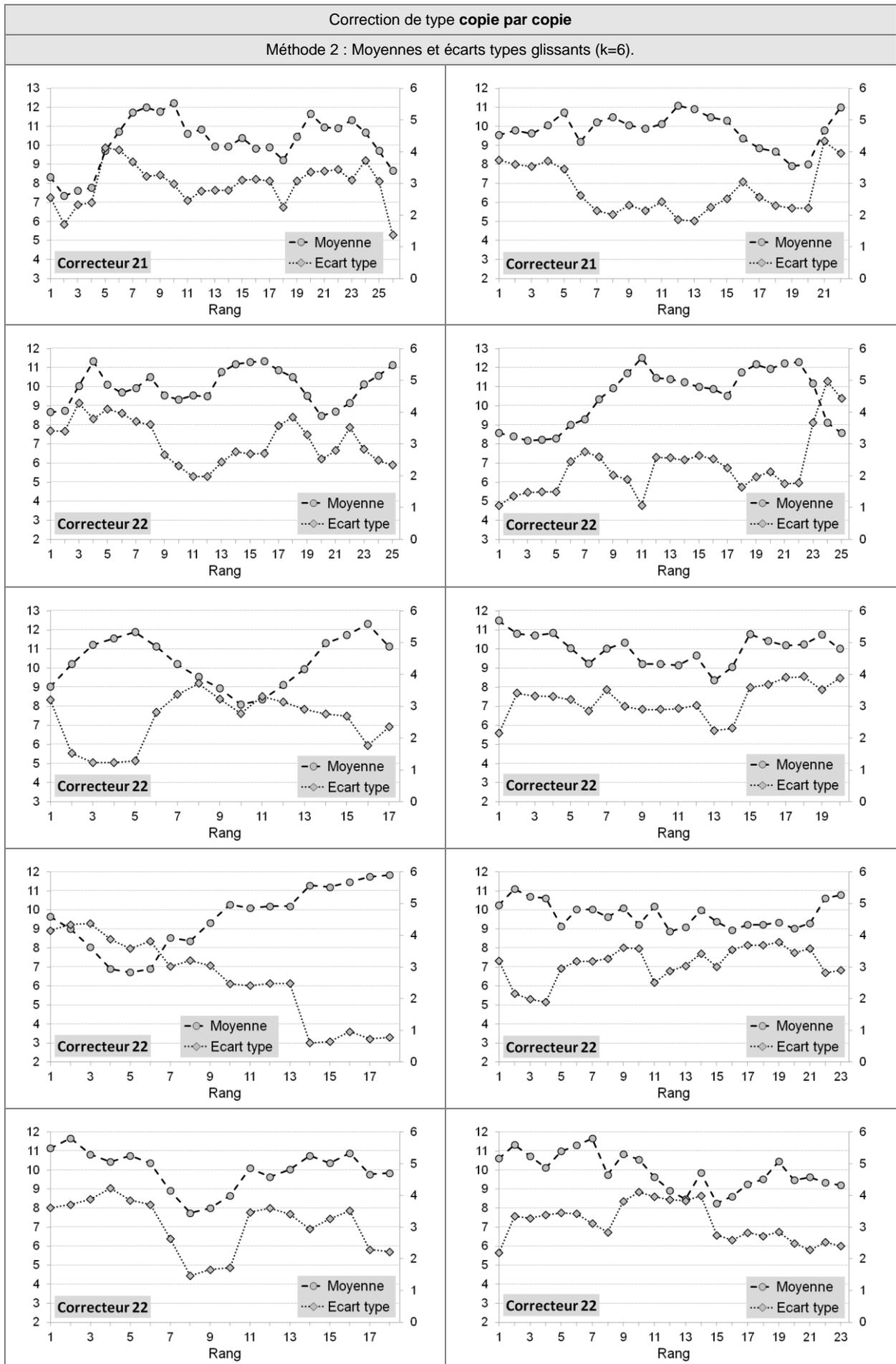


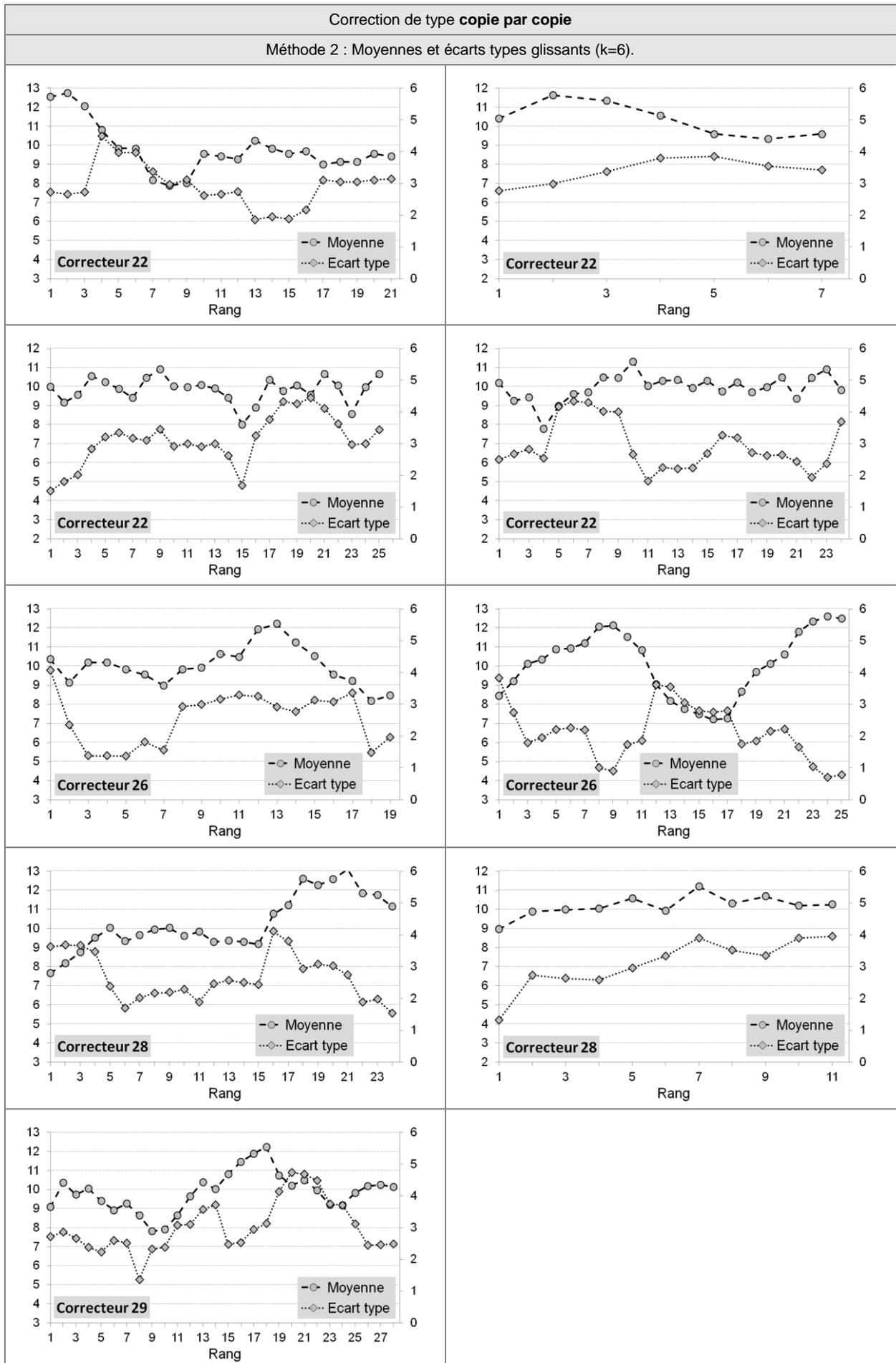




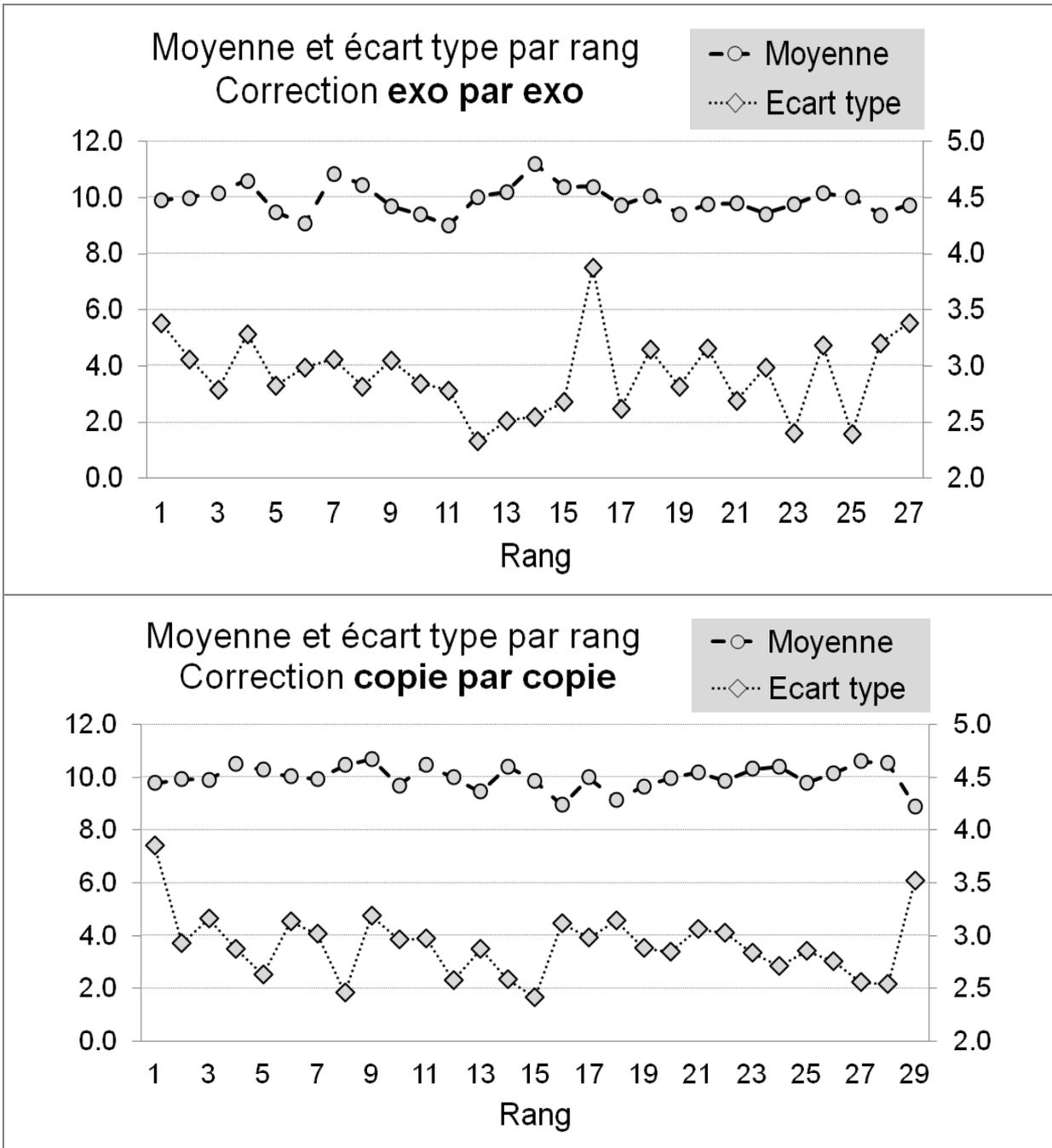








Annexe 5: Résultats obtenus par la méthode 3 : moyennes et écarts types par rang.



Influence du rang de correction dans l'évaluation en mathématiques sur une échelle de 21 degrés (0-20).

Une étude simultanée de la position et de la distribution des notes de 94 séries de copies.

Auteur : Emmanuelle CANCEZ

Directeur de mémoire : Yvan Abernot

Année : 2016

Nombre de page : 30 pages hors Annexes

Résumé :

Cette recherche en docimologie s'intéresse à l'influence du rang de correction d'une copie lors de la correction d'une série de copie sur une échelle de 21 degrés, et ce, plus particulièrement dans l'évaluation des mathématiques. L'idée novatrice développée dans ce projet de recherche est d'étudier simultanément l'évolution de la moyenne et l'évolution de l'écart type (des notes) en fonction du rang de correction des copies. La principale hypothèse de travail autour de laquelle s'articule la problématique est l'hypothèse selon laquelle « au fur et à mesure de la correction la moyenne diminue et l'écart type augmente ». Au total, 94 séries de notes issues d'évaluations différentes ont fait l'objet d'une analyse afin de répondre aux différentes questions de recherche. Les effets d'ordres tels qu'ils ont été décrits dans le passé, n'ont pas été mis en évidence de façon claire. Bien que la principale hypothèse de travail n'ait pu être confirmée, elle ne peut être infirmée. En effet, les données expérimentales recueillies et analysées sont issues de corrections « réelles » ayant de ce fait pu être impactées à la fois par des effets parasites tels que les effets d'ordre, mais également par d'autres effets influant sur la validité d'une évaluation, comme par exemple, ceux de contamination, de stéréotypie et de halo.

Mots clés : docimologie, validité de l'évaluation, effets d'ordre, mathématiques

Summary :

This docimological study deals with the influence of marking rank during the process of marking a set of copies and, more precisely, in the frame of mathematics evaluation on a 21 degree scale. The innovative idea developed in the frame of this research project is studying simultaneously the evolution of the following two parameters: average and standard deviation, with marking rank. The main working hypothesis around which problematic has been set up is "average values decrease and standard deviation values increase while the marking process goes on". A total of 94 series of marks coming from different evaluations have been analyzed in order to answer the questions asked in accordance with the problematic. Marking rank effects, as described in the past by previous studies, have not been highlighted in the current study. Although the main working hypothesis has not been corroborated, it cannot be rejected. Indeed, experimental data which have been collected and analyzed in the frame of the current study, comes from "actual" marking situations and as a consequence may have suffered others unwanted and/or side effects as marking rank effect but also other effects affecting measure validity as, for example, contamination effect, stereotype effect and halo effect.

Keys words : docimology, measure validity, ranking effects, mathematics