

Le 2000 mètres au professorat des écoles



Constats et propositions d'amélioration de l'épreuve et des barèmes

Jean-Pierre Cleuziou et Claire Pontais

Avec la collaboration de formateurs en IUFM et de conseillers pédagogiques évaluateurs

Juillet – Septembre 2003

L'essentiel de ce rapport a été établi pour le 1^{er} juillet 2003. Transmis en juillet au ministère et aux enseignants collaborateurs. Il a intégré, en septembre, les remarques de ces enseignants et quelques modifications non connues en juillet

Sommaire

Les raisons de cette étude	3
Introduction	4
1. Analyse critique de l'épreuve	5
1.1. Rappel sur l'épreuve actuelle	5
1.2. Interprétation et commentaires	5
1.3. Les critiques des enseignants	6
1.3. Quelles évolutions souhaitables et possibles	6
2. Méthode d'étude des performances et d'élaboration des barèmes	7
2.1. Le recueil des performances et des informations	7
2.2. Le choix d'une échelle de notation de 0 à 20	7
2.3. La relation notes - performances	8
2.4. Transposition sur 15 et 5 points	8
3. CRPE femmes : les temps sur 2000 m	9
3.1. Étude des performances	9
3.2. Recherche du barème et conséquences sur les notes	9
3.3. Barème proposé	10
3.4. Évolutions possibles, variantes	10
4. CRPE hommes : les temps sur 2000 m	11
4.1. Étude des performances	11
4.2. Recherche du barème et conséquences sur les notes	11
4.3. Barème proposé et distribution des notes	12
4.4. Évolutions possibles, variantes	12
5. CRPE femmes : les écarts au projet	12
5.1. L'étude des performances	12
5.2. Recherche d'un barème	13
5.3. Relation entre la performance et l'écart au projet.....	14
6. CRPE hommes : les écarts au projet	15
6.1. L'étude des écarts au projet	15
6.2. Recherche d'un barème	15
6.3. Relation entre la performance et l'écart au projet.....	16
7. Les écarts au projet : barème mixte ?	16
Conclusions et résumé des propositions	17
Propositions	17
Pourquoi	17
Remarque.....	18
Bibliographie sommaire	19
Annexe 1 : comparaison avec d'autres résultats	19
Comparaison avec les tables du test de Cooper	19
Comparaison avec des résultats de lycéens	20
Annexe 2 : bref mémento statistique	21
1. Construire l'échantillon, recueillir les données	21
2. Décrire et résumer les résultats.....	21
3. Estimer la population parente.....	22



Les raisons de cette étude

En novembre 2002, les dispositions initiales concernant la réglementation et les barèmes pour le 2000 mètres au CREP (concours de recrutement au professorat des écoles) ont provoqué de vives critiques de la part des collègues chargés de la préparation des candidats et de notre commission pédagogique. Nous avons organisé dans l'urgence une consultation, fait des propositions, et certaines améliorations ont pu être obtenues. Nous avons demandé aussi que pour cette épreuve, comme pour les autres épreuves d'examen et de concours, soient réalisées des études approfondies sur le niveau réel des candidats.

L'administration ne voulant pas s'engager sur ce plan, le SNEP a sollicité le Centre EPS et Société pour la réalisation de cette étude. Claire Pontais et Jean-Pierre Cleuziou se sont associés pour développer cette étude en compagnie des enseignants d'IUFM intéressés. Claire Pontais, titulaire d'un DEA Sciences de l'éducation et didactiques, est notre responsable nationale des IUFM (formation des enseignants du premier degré). JP Cleuziou a notamment réalisé en 1994 une thèse sur les méthodes d'élaboration de tables de cotation et a développé plusieurs recherches sur l'évaluation en EPS ; il est membre du GEDIAPS, équipe de recherche STAPS de l'université de Paris XII.

Le déroulement des épreuves lors de la session de juin 2003 a confirmé nos craintes et le bien-fondé de nos critiques et de celles des formateurs. Le présent rapport fait la synthèse de ces remarques et analyse les performances de plus de 1000 candidats pour déboucher sur des propositions d'adaptation de l'épreuve et des barèmes.

Au-delà de cette contribution, le SNEP et le Centre EPS et Société veulent montrer que l'étude des performances des collégiens, des lycéens et des étudiants est non seulement nécessaire mais possible pour concevoir des épreuves d'examens et de concours les plus aptes à révéler les savoirs et les compétences acquis par les populations évaluées. Ce problème se retrouve notamment au baccalauréat mais aussi au CAPEPS.

**Alain Becker, secrétaire national pédagogique du SNEP,
juillet 2003**

Introduction

Au cours de la session de juin 2003, les candidats(es) au CRPE avaient le choix entre trois épreuves définies par des dispositions nationales : une épreuve de danse, une épreuve de badminton et une épreuve de 2000 mètres¹.

Définir des modalités d'évaluation équitables pour ces trois épreuves est une tâche particulièrement difficile car aucune étude sur le niveau de prestation des candidats(es) dans ces épreuves n'existe à l'heure actuelle. Les épreuves sont nouvelles et les aptitudes sportives de la population particulière que constituent les candidats(es) au CRPE sont mal connues, même si les formateurs ont souvent une connaissance intuitive des niveaux. Il s'agit d'une population fortement féminisée, très diverse du point de vue des pratiques sportives mais aussi des âges. Beaucoup de candidats(es) n'ont eu aucune pratique sportive pendant plusieurs années après avoir quitté le lycée.

Les premières dispositions concernant l'évaluation dans ces trois épreuves ne pouvaient donc résulter que d'une estimation approximative du niveau des candidats(es).

Début novembre 2002, la parution des premiers barèmes pour le 2000 m au CRPE a suscité chez les formateurs IUFM de vives critiques, tant les modalités de notation paraissaient inadaptées pour l'évaluation de la « performance » (temps réalisé) comme pour celle de la maîtrise d'un projet (écart entre le temps prévu et le temps réalisé)².

Les remarques qui ont alors été faites ont amené, dans l'urgence, une adaptation des barèmes initiaux sur la base d'estimations nécessairement approximatives. Par ailleurs l'épreuve est restée très discutée dans sa réglementation.

Le déroulement des épreuves a confirmé les craintes des enseignants sur trois plans : réglementation de l'épreuve, barèmes, correspondance des exigences entre les trois épreuves : danse, badminton et 2000 mètres. Les notes de badminton, de danse et d'entretien s'avèrent nettement meilleures que celles du 2000 m et ce n'est que dans cette épreuve qu'on constate un grand nombre de 0, ces notes nulles ne concernant pratiquement que les femmes³.

¹ Pour cette épreuve, la note tient compte du temps réalisé mais aussi de l'écart à la performance visée (écart au projet).

² Les premiers barèmes publiés au BO prévoyaient les cotations suivantes (résumé). On peut vérifier que cette cotation se fonde sur la table Letessier.

Notes >	0	5	10	15	20
Filles	11.00	9.48	8.45	7.49	7.00
Garçons	9.15	6.23	7.09	8.16	5.44

Toute performance au-delà de 11 mn pour les filles et de 9mn15 pour les garçons était notée 0/20. Il était prévisible que plus de 30% des filles obtiendraient la note 0 en performance et ne seraient notées que pour l'écart au projet. Un tel barème désavantageait de plus très fortement les filles ce qui n'était pas admissible dans un concours.

³ Selon les bilans départementaux que nous avons pu consulter en septembre 2003, on constate 6 à 11% de notes nulles pour l'épreuve de 2000 m alors que les notes nulles sont quasi inexistantes en

La présente étude est centrée sur le 2000 mètres parce que cette épreuve est largement choisie et parce qu'elle pose des problèmes qu'il est urgent de résoudre.

La majorité des candidats(es) choisissent cette épreuve. On sait qu'une épreuve barémée permet aux candidats d'éviter les surprises de l'évaluation. Le 2000 m est une épreuve où les candidats peuvent réaliser un entraînement personnel sans installations sophistiquées et progresser nettement. C'est aussi une épreuve à la portée de presque tous après entraînement et elle pose des problèmes intéressants de gestion des efforts. Enfin c'est une épreuve facile à organiser.

L'épreuve est donc intéressante mais il est urgent de tenter de l'améliorer. L'analyse des performances (temps et écart au projet) révèle que sur 837 candidates, 6.1% d'entre elles ont obtenu la note 0 (leurs niveaux n'a donc pu être différencié) alors que cette note n'a été attribuée qu'à 1% des 198 hommes dont nous avons eu les résultats. Il faut ajouter qu'au vu des performances, les candidates ont obtenu en moyenne des notes inférieures de près de 2 points par rapport aux candidats.

Pour tenter d'améliorer cette situation, nous partirons de l'analyse de l'épreuve et ferons part des remarques de formateurs et d'évaluateurs.

Après avoir apporté quelques précisions sur les méthodes utilisées, nous étudierons les temps des candidats(es) et leurs écarts au projet pour déboucher sur des propositions de barèmes.

Enfin en conclusion nous synthétiserons cet ensemble à travers une proposition de réglementation de l'épreuve accompagnée de barèmes.

Il est donc possible de lire d'abord la conclusion pour remonter ensuite aux développements qui la précèdent et justifient les propositions.

Les problèmes que s'efforce de résoudre cette étude ne sont pas isolés, on les retrouve notamment au baccalauréat mais aussi au CAPEPS.

À travers cette étude nous espérons aussi montrer plus généralement l'intérêt d'une analyse méthodique des épreuves et des performances pour déboucher sur des solutions adaptées aux populations à former. Dans une telle perspective les enseignants ont un rôle de premier ordre à jouer.

Pour l'évolution de l'EPS et la formation continue des enseignants on a tout intérêt à travailler dans ce sens.

badminton et en danse. Les moyennes en badminton et en danse sont souvent supérieures de 2 à 3 points à celle du 2000 m.

1. Analyse critique de l'épreuve

1.1. Rappel sur l'épreuve actuelle

Elle consiste en un 2000 mètres pour les femmes et les hommes. Avant le départ de l'épreuve, chaque candidat annonce le temps qu'il vise. Les candidats sont alors évalués sur 15 points pour la performance (temps à l'arrivée) et pour 5 points pour l'écart au projet.

Pour l'évaluation de la performance, les barèmes sont différenciés selon le sexe (sur 15 points au demi-point près).

Pour l'écart au projet (appelé aussi « bonification »), le barème est le même pour les hommes et pour les femmes (sur 5 points au point près). Au total alors que la notation se fait au demi-point de 0 à 15, elle se fait au point près de 15 à 20. Cette discontinuité dans la notation peut paraître surprenante.

On remarque que le barème performance femmes est pour l'essentiel parallèle à celui des hommes avec un simple décalage de 2 minutes.

Notes	Perf (F)	Perf (H)	Différence
15	8.50	6.50	2.00
12	9.30	7.22	2.08
9	10.16	7.50	2.26
7.5	10.43	8.40	2.03
6	11.10	9.10	2.00
3	12.10	10.10	2.00
0	>13	>11	2

Il faut encore noter deux dispositions très importantes.

- Les candidats ne peuvent pas courir avec une montre ou un chronomètre et ils ne reçoivent aucune indication de temps de passage durant les 2000 m.
- Le temps projeté ne peut être hors barème : une femme qui vaut 16 minutes sur 2000 mètres ne peut donc avoir comme projet que 13 minutes, ce qui entraînera automatiquement la note globale 0 puisque au-dessus d'un écart de 30 secondes, l'écart au projet est noté 0.

1.2. Interprétation et commentaires

Il apparaît clairement que les concepteurs de l'épreuve se sont efforcés de concilier deux exigences. Les candidats doivent réaliser une performance athlétique (réaliser le meilleur temps) mais en même temps ils doivent démontrer un savoir méthodologique référé à la pratique de l'athlétisme : tout athlète éduqué est en effet capable de viser une performance et de la réaliser avec une certaine marge d'erreur. De plus la condition d'un entraînement efficace en demi-fond repose sur cette capacité de courir à des allures maîtrisées.

La performance comptant trois fois plus que la maîtrise du projet, cette épreuve conserve le sens principal d'une activité de demi-fond.

La prise en compte de l'écart au projet est intéressante dans son principe pour plusieurs raisons :

- Elle repose sur une capacité d'auto-évaluation en action qui doit être développée en EPS (élaboration d'un projet réaliste et maîtrise de sa réalisation).
- Elle prend ces racines dans une activité authentique des coureurs de demi-fond.
- Elle est indépendante des facteurs énergétiques (la corrélation entre cette activité hautement cognitive et la performance est proche de zéro). On évalue donc une capacité différente de la puissance maximale aérobie qui, elle, conditionne principalement le temps sur 2000 mètres. Les deux évaluations ne sont pas redondantes.
- Enfin il faut souligner que la procédure d'évaluation est simple, il suffit de faire une soustraction entre le projet et le temps réalisé.

Tout cela tend à justifier cette épreuve. Encore faut-il que les barèmes de notation et les règles de déroulement soient bien adaptés. C'est à ce niveau que l'épreuve s'avère le plus discutable.

Avant d'examiner les critiques des formateurs, nous ferons encore trois remarques.

- La recherche simultanée de la meilleure performance et de la réalisation précise d'un projet peut s'avérer en partie contradictoire : le coureur qui s'est sous-estimé est capable de faire mieux que son temps annoncé. S'il se freine il est en contradiction avec le sens de l'activité demi-fond, s'il accélère, sa note écart au projet diminue. On peut rétorquer que le candidat n'a pas su estimer ses réelles possibilités et qu'il est juste que cela soit sanctionné.
- La réalisation précise d'un projet dépend fortement des conditions d'information durant la course. Elle dépend aussi de l'intensité de l'effort et des risques pris ainsi que des conditions atmosphériques. Quand un athlète recherche un temps de 5% inférieur à son record, sans vent et qu'il bénéficie de temps de passage tous les tours, il est assez aisé de réaliser son projet à moins de 5 secondes près. En revanche la tâche est très complexe s'il court par fort vent et sans aucune indication de temps de passage. Pour évaluer un écart au projet il est donc essentiel de préciser les conditions d'information.
- Soulignons aussi l'une des difficultés rencontrées dans ce type d'épreuve. Un candidat qui marcherait régulièrement pendant les 2000 m, obtiendrait 0/15 pour la « performance » mais 5/5 pour l'écart au projet. Tous les candidats étant capables de marcher 2 km, est-il acceptable que la note minima d'un candidat soit automatiquement 5 ? Par ailleurs pour des impératifs de déroulement, il est nécessaire de borner en temps la durée de l'épreuve. Pour ces deux raisons, il n'est donc pas aberrant de fixer une limite supérieure de temps pour l'épreuve en tenant bien compte du niveau des candidats.

1.3. Les critiques des enseignants

a) Exemples de remarques reçues

Nous citons quelques témoignages révélateurs des critiques portées à l'épreuve.

● Premier enseignant : « 13 filles et 2 garçons ont dû annoncer le meilleur temps permis alors qu'ils réalisaient des temps inférieurs. Ils ont donc perdu des points de contrat (la meilleure fille fait 8'10 et donc a 0 de contrat) ARCHI NUL ! Je me suis fendu d'une lettre adressée au rectorat remettant en cause cette aberration... Elle a été signée par tous les CPC et CPD présents ».

● Un autre : « ...Plusieurs filles qui étaient hors barème se sont vues pénalisées par leur perf, c'est tout de même le comble !! Nous avons la consigne de ne communiquer aucun temps, tout le monde a trouvé ça débile mais on a tous eu la même attitude, on a respecté les décisions (académiques ?) pour que tous les candidats soient logés à la même enseigne. Cela dit, les plus malins avaient des chronomètres planqués... Nous avons consigné nos remarques par écrit, soulignant le manque de sens et l'absurdité du barème qui pénalise les plus performants... Ce serait dommage de virer l'épreuve mais il faut absolument la modifier. Le contrat devrait, à la fois valoriser les efforts des moins performants et "récompenser" les bonnes perf... Ça devrait pouvoir se faire. »

● Un troisième : « Il faut dire, qu'en particulier chez les filles, certains candidats étaient partis de très bas. Les résultats sont donc corrects par rapport au niveau de départ. Par contre des problèmes que nous avons signalés auprès des responsables académiques des épreuves, se sont confirmés.

- Équité entre les épreuves : personne ne peut avoir 0 en maîtrise d'exécution en Bad ou en Danse quel que soit son niveau. Or des candidats(tes) ont eu zéro à l'épreuve du 2000 m alors qu'ils ont effectué le temps annoncé : 14 mn pour une candidate en surcharge pondérale qui a par ailleurs progressé de 5 mn depuis le début de l'année !!!...)

- Équité au sein de l'épreuve même, d'un centre à l'autre. Une chance que la tramontane n'ait pas soufflé ce jour-là, car impossible d'assurer un contrat aussi "pointilleux" (les entraînements ont mis en évidence les jours de vent, sur des perfs "stabilisées", des décalages de plus de 20 secondes...).

- Enfin, nous sommes toujours un grand nombre à être persuadés que la natation aurait bien plus sa place dans le cadre de ce concours au regard du futur métier de ces candidats. »

● Signalons aussi un avis (oral) divergent : « si l'on donne les temps tour par tour, le problème de l'éducation à la maîtrise des allures est escamoté. Un candidat qui s'est peu préparé peut réaliser son projet avec une marge d'erreur faible si on lui fournit les temps de passage tour par tour. Or on vise une activité fine de contrôle de soi. Il faut donc que les conditions de déroulement soient exigeantes sinon on n'éduque personne ». Et ce collègue ajoute : « Il ne faut pas supprimer cette épreuve, il faut la conserver et l'améliorer. Je suis pour une épreuve de natation en plus ».

b) Synthèse des principales critiques

- Les barèmes de performances ne sont pas bien adaptés notamment pour les filles, certaines sont capables de faire mieux que la cote supérieure, mais c'est surtout au niveau

inférieur que cela ne va pas. Trop de filles sont hors barème. Certains enseignants ajoutent que les barèmes ne sont pas trop exigeants mais sans préciser à quels niveaux de la cotation ils situent cette appréciation.

- L'évaluation de la maîtrise du projet ne semble pas remise en cause dans son principe, mais il faut en améliorer les conditions.
- L'information durant l'épreuve (temps de passage interdits) pose problème. (Rappelons qu'en demi-fond les officiels sont autorisés à donner des temps de passage). Personne ne critique l'interdiction de chronomètre personnel, mais un avis majoritaire semble se dessiner pour que soient donnés les temps de passage tous les 400 mètres.
- Certains enseignants aimeraient que cette épreuve soit une course de longue durée qu'il faudrait définir. Mais il semble qu'un accord peut se faire sur la conservation de cette épreuve à condition de l'améliorer. Par ailleurs il serait souhaitable d'ajouter une quatrième épreuve (natation).
- Enfin il semble qu'il faille mieux ajuster le niveau des exigences dans les différentes épreuves (danse, badminton et 2000 mètres). Il est anormal qu'au 2000 m les zéros soient assez fréquents alors que cela ne l'est pas en danse et en badminton.

1.3. Quelles évolutions souhaitables et possibles

- Il faut d'abord adapter les barèmes en tenant compte du niveau des candidats(es) : on verra plus loin que le niveau inférieur de la cotation devrait être proche de 11 minutes pour les garçons et de 16 minutes pour les filles.
- Pour des raisons de faisabilité et de validité de l'épreuve, il faut fixer une durée limite de l'épreuve en tenant compte d'une révision des barèmes. Par exemple :
 - Hommes : entre 11 et 15 minutes, les points d'écart au projet seraient diminués au pro rata du temps. Au-delà de 15 minutes le candidat aurait alors 0.
 - Femmes : entre 16 et 20 minutes, les points d'écart au projet pourraient être diminués au pro rata du temps. Au-delà de 20 minutes la candidate aurait 0.
- Il faut faire disparaître la limite vers le haut. On verra que l'étude des performances conduit à proposer de « durcir » le barème pour les meilleurs candidats(es) ce qui permet de mieux distinguer les candidats les plus forts.
- Actuellement le terme « bonification » qui est attribué à la notation de l'écart au projet est impropre puisque la notation de l'épreuve est sur 20 avec un décompte 15 + 5. Il n'y a donc pas de bonus. Pour conserver l'idée intéressante de bonus, il serait intéressant de noter la performance sur 16 et l'écart au projet sur 5, toute note supérieure à 20 étant ramenée bien sûr à 20. Cela permettrait de valoriser la « performance » pour les meilleurs sans diminuer la part attribuée à la maîtrise du projet.

2. Méthode d'étude des performances et d'élaboration des barèmes

Étalonner un barème consiste à définir des correspondances entre des notes et des performances sur la base de conventions précises qui donnent un sens aux notes. Cette opération suppose

- le choix d'une échelle de notes (conventionnelle),
- l'étude d'un échantillon de performances représentatif,
- le choix d'une méthode de mise en relation des notes et des performances.

Dans ce qui suit, le terme générique « performance » s'applique au temps réalisé à l'arrivée du 2000 mètres mais aussi à l'écart au projet. Ces deux types de résultats constituent des performances au sens large du terme.

2.1. Le recueil des performances et des informations

Ces informations ont bien sûr été collectées en respectant un strict anonymat : sexe, âge (si possible), temps prévu, temps réalisé, niveau d'entraînement individuel (si possible).

a) Les performances

Nous avons recueilli les résultats de 837 filles et de 198 garçons sans aucune sélection en situation d'évaluation réelle au CRPE. Réalisées par des candidats motivés, ces performances peuvent être considérées comme fiables.

Toutefois les conditions de réglementation des épreuves ont introduit des biais : impossibilité de se fixer un projet hors barème vers le haut et vers le bas ce qui a influencé le comportement de certains candidats.

Malgré ces réserves, on peut considérer que nous disposons d'échantillons représentatifs d'effectif important pour les filles et d'effectif assez important pour les garçons. Une estimation assez fiable du niveau des candidats peut donc être faite.

b) Données complémentaires

Pour interpréter ces résultats nous avons souhaité recueillir des informations sur :

- l'âge des candidats
- la préparation individuelle des candidats selon une grille d'appréciation en 5 points.

Les informations sur la préparation des candidats ont été trop peu nombreuses pour en tirer des enseignements.

Nous avons obtenu l'âge de 543 candidats(es). Les candidates (440) sont âgées de 20 à 45 ans (médiane : 26 ans), les hommes (103) de 21 à 41 ans (médiane : 27 ans).

âge	Femmes		Hommes	
20-25 ans	191	43%	37	36%
26-30 ans	147	33%	52	50%
31-35 ans	55	13%	10	10%
36-40 ans	39	9%	3	3%
41-45 ans	8	2%	1	1%

Le pourcentage des candidates âgées de plus de 30 ans est nettement plus important que celui des hommes (24% et 14%).

On constate que la corrélation entre âges et performances sur 2000 m est faible ($r = 0.12$ pour les femmes et 0.19). Elle est nulle entre âge et écart au projet.⁴

Nous avons donc centré l'étude sur l'élaboration des barèmes indépendamment de l'âge.

2.2. Le choix d'une échelle de notation de 0 à 20

Ce sont les conventions de notation qui donnent un sens aux notes et ces conventions doivent concourir au mieux à l'atteinte des objectifs recherchés (renforcer la motivation pour se préparer, distinguer les candidats dans un concours...).

Au moment de l'étalonnage, il est courant de tendre vers une distribution gaussienne des notes. Il suffit alors de convenir de la moyenne et de l'écart type des notes pour que chaque note soit définie en termes d'écart à la moyenne et de fréquence théorique. Voici quelques exemples qui montrent l'importance des choix sur la moyenne et l'écart type ; ce sont ces choix qui rendent plus ou moins sélectif l'échelle des notes.

	si $m = 10$, $\sigma = 3$	si $m = 10$, $\sigma = 4$	si $m = 12$, $\sigma = 3$	si $m = 12$, $\sigma = 4$
10	50.0%	50.0%	74.8%	69.1%
15	4.8%	10.6%	15.9%	22.7%
% de notes supérieures à 10 et 15 selon m et σ				

Remarquons que pour un concours le choix d'une distribution gaussienne des notes peut être contesté car une telle distribution distingue peu les sujets moyens. On peut chercher à augmenter la dispersion des notes des sujets moyens mais ce sont alors les sujets faibles et/ou forts qui sont moins bien différenciés, et l'on a dilaté artificiellement les écarts de notes au centre alors que les sujets sont en réalité très proches les uns des autres.

Nous avons donc choisi la solution la plus classique et la plus commode : tendre vers une distribution gaussienne. Par ailleurs, nous avons tenu compte de quelques indications :

- Il s'agit d'un concours donc le barème doit différencier les candidats.
- Le barème doit être incitatif à tous les niveaux de performances.

⁴ Si le nombre de performances est trop faible pour en tirer des conclusions indiscutables, on note toutefois une baisse du niveau de performance avec l'âge.

	Femmes				Hommes			
âges	20-25	26-30	31-35	35-45	20-25	26-30	31-35	35-41
Nombre	185	142	54	46	36	52	10	3
Moyenne	10.43	10.47	11.07	11.07	8.02	8.09	8.44	8.28

- La note 10 ne peut être accessible à tous. Nous ne sommes pas en EPS ni au baccalauréat.

Faute de directives précises, il a été choisi pour construire les barèmes, de tendre vers une moyenne de 10 et un écart type de 3.333. Conformément aux propriétés de la loi normale :

- 99.7% des candidats seront compris entre les notes 0 et 20 ;
- 50% des candidats auront 10 ou plus ;
- 93% des notes seront comprises entre 5 et 15.

Remarque :

Les barèmes ayant été calculés en fonction de ces choix, il sera facile de les convertir si l'on décide d'une autre échelle de notes. Par exemple moyenne = 10 et écart type = 4, ce qui permet de mieux différencier la majorité des candidats, les extrêmes, peu nombreux, seront un peu moins différenciés.

2.3. La relation notes - performances

Il ne suffit pas d'avoir adopté une échelle de notes, de disposer d'échantillons de performances et d'informations issues d'autres études, il faut encore une méthode pour construire la relation notes – performances. Cette méthode doit permettre d'intégrer au mieux toutes les informations disponibles.

Nous allons voir que nous ne pouvons pas utiliser les deux méthodes classiques d'étalonnage : selon les notes centrées réduites ou selon le centilage.

a) Brève critique des deux méthodes classiques d'étalonnage

Ces deux méthodes s'efforcent généralement de tendre vers une distribution gaussienne des notes.

La méthode des notes centrées réduites (relation linéaire entre les notes et les performances)⁵ est une méthode très simple mais elle ne convient que lorsque les performances s'écartent peu d'une distribution gaussienne, or nous verrons que les distributions des temps sur 2000 mètres et des écarts au projet sont très éloignées d'une distribution gaussienne.

La méthode du centilage⁶ convient quelle que soit la forme de la distribution des performances mais elle épouse de très près les caractéristiques de l'échantillon : les barèmes subissent donc les conséquences des variations aléatoires notamment dans les extrêmes. Par ailleurs on ne débouche pas avec cette méthode sur des barèmes paramétrés et il est impossible d'agir sur un petit nombre de repères pour adapter le barème en tenant compte d'autres sources de connaissances.

b) Méthode utilisée

La méthode que nous utiliserons permet de tenir compte des asymétries très variables des distributions des performances et de déboucher sur des barèmes paramétrés à l'aide de trois performances repères P1, P2 et P3. Connaissant ces trois

paramètres il est possible de calculer l'ensemble du barème et de l'adapter en agissant sur ces paramètres.

Il n'est pas possible de décrire ici avec précision les fondements de cette méthode. Nous la résumerons donc par ces quelques indications.

- Au moment de l'étalonnage, nous choisissons de tendre vers une distribution gaussienne des notes (voir plus haut).
- La relation mathématique entre notes et performances doit alors satisfaire à l'exigence suivante : partant de distributions de performances aux formes très variées, il s'agit de tendre toujours vers une distribution gaussienne des notes. Cela est le plus souvent obtenu lorsqu'on obtient une distribution des notes symétrique (asymétrie nulle).
- Pour cela la relation entre les notes (n) et les performances (p) est du type $n = a \cdot \ln(p - P_0) + b$ dans laquelle P_0 est un paramètre qui permet de faire varier l'asymétrie des notes. On fait donc varier P_0 selon un programme informatique jusqu'à ce que l'asymétrie des notes devienne pratiquement nulle. On obtient alors une distribution des notes proche d'une distribution gaussienne et l'on peut en déduire les paramètres P1, P2 et P3 qui résument le barème et permettent de le reconstruire à la précision souhaitée.
- Connaissant P1, P2 et P3 il est aussi possible d'ajuster le barème en agissant sur ces paramètres, cela s'avère utile notamment quand l'effectif de l'échantillon n'est pas bien représentatif et suffisamment fourni.
- En agissant sur ces trois paramètres on peut enfin adopter d'autres conventions sur la forme de la distribution.

Le traitement des résultats permettra d'illustrer ces quelques repères.

2.4. Transposition sur 15 et 5 points

L'étude des barèmes se fera sur 20 points ce qui permet de définir des niveaux d'exigence comparables pour les temps sur 2000 mètres et pour les écarts au projet .

Dans la conclusion une transposition sur 15 et 5 points sera proposée pour respecter la présentation actuelle de la notation.

⁵ La moyenne des notes correspond à la moyenne des performances et lorsque les performances varie d'un écart type les notes varient d'un écart type.

⁶ La méthode du centilage est fondée sur la correspondance des pourcentages. Ainsi la médiane des performances (et non la moyenne) correspond à la médiane des notes ; la performance dépassée par 10% des sujets correspond à la note dépassée par 10% des sujets, etc.. Dès qu'on a convenu d'une échelle de notes définies en termes de pourcentages, la correspondance des performances est facile à calculée avec un tableur (fonction centile).

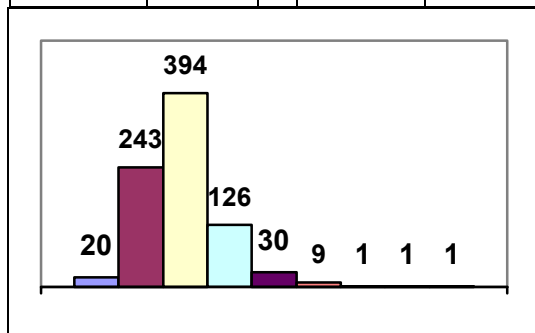
3. CRPE femmes : les temps sur 2000 m

Rappelons que les règles de l'épreuve ne permettaient pas à une candidate de niveau faible (réalisant plus de 13mn30s) d'obtenir une note de performance et d'écart au projet supérieure à zéro : 6.1% des 837 candidates dont les performances nous sont parvenues ont obtenu la note zéro.

Ces conditions ont aussi pu provoquer des abandons ou des performances très faibles (course sans motivation). Par ailleurs le barème tendait à limiter l'expression des meilleures candidates (voir ci-dessus 1.). Pour les meilleures comme pour les plus faibles l'échantillon est donc biaisé.

3.1. Étude des performances

Résumé A		Lim. Sup.	Effectifs
Effectif	825	8.36	20
Moyenne	10.483	10.09	243
Médiane	10.370	11.42	394
Ec type	1.224	13.15	126
Asymétrie =	1.412	14.48	30
mini	7.03	16.21	9
maxi	21.00	17.54	1
		19.27	1
		21.00	1



L'effectif est important (825 temps + 12 abandons) ce qui permet d'avoir une idée précise du niveau des candidates (sauf aux extrémités compte tenu de la réglementation).

Les performances sont très dispersées, de 7mn03 à 21mn00. De plus il faut noter 12 abandons (1 seul pour les garçons). Trois performances sont excellentes : 7mn03, 7mn29 et 7m32 puis 7mn55, 7mn56, 7mn58...

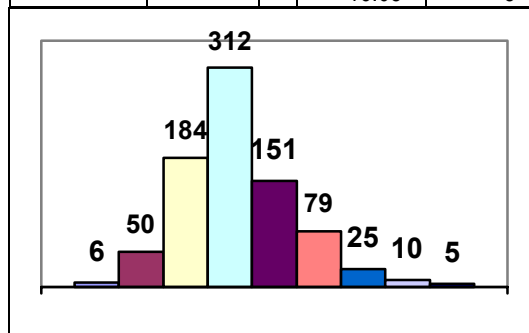
L'asymétrie des performances est très forte (histogramme, coefficient d'asymétrie, médiane différente de moyenne).

Il faut toutefois noter que cette asymétrie est fortement liée à trois performances nettement plus faibles que les autres.

Si l'on ne tient pas compte de ces trois performances, on obtient le bilan suivant.

Nous rechercherons un barème sur la base de ces 822 performances.

Résumé B		Lim. Sup.	Effectifs
Effectif	822	8.04	6
Moyenne	10.465	9.04	50
Médiane	10.370	10.05	184
Ec type	1.170	11.05	312
Asymétrie	0.796	12.06	151
mini	7.03	13.06	79
maxi	16.08	14.07	25
		15.07	10
		16.08	5



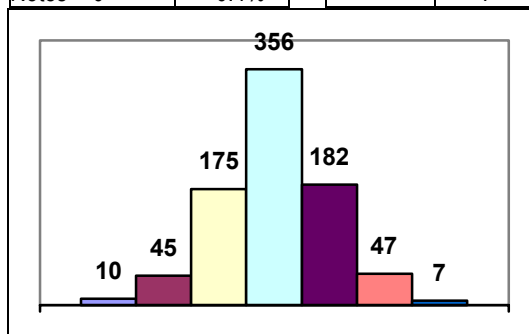
3.2. Recherche du barème et conséquences sur les notes

Si nous recherchons un barème qui permette de normaliser les notes nous parvenons à la solution suivante (résumée) :

Notes	0	5	10	15	20
Perf	15.42	12.49	10.39	9.03	7.51
	P1		P2		P3

Le bilan des notes s'établit ainsi.

Moyenne	10.00	(note page suivante)	
Ec type	3.33	Lim. sup.	Effectif
Asymétrie	0.002	m - 2.5 σ	10
Mini =	-0.6	m - 1.5 σ	45
Maxi =	24.4	m - 0.5 σ	175
Médiane	10.10	m + 0.5 σ	356
Notes >= 20	0.4%	m + 1.5 σ	182
Notes >= 10	51.5%	m + 2.5 σ	47
Notes < 0	0.4%		7



Note sur la répartition en classes des notes :

Pour vérifier si une distribution est gaussienne, on adopte habituellement des limites en fonction de la moyenne (m) et de l'écart type (σ), avec un nombre impair de classes pour vérifier que la classe centrale est modale.

Ici on a retenu 7 classes dont on a indiqué seulement la limite supérieure (la limite inférieure étant la limite supérieure de la classe précédente). La première classe n'a donc pas de limite inférieure, la dernière pas de limite supérieure.

Discussion sur P3 (note 20) : Le barème propose 7mn51, c'est nettement plus exigeant que les 8mn50 du barème CRPE 2003 mais on observe six temps inférieurs à 8 minutes. Nous retiendrons P1 = 8.00 pour donner un repère simple.

Discussion sur P2 (note 10) : le barème propose 10mn39. Cette proposition paraît bien en accord avec le niveau médian des candidates, le barème CRPE 2003 indique pour sa part 10.43. La convergence est évidente. Nous retiendrons P2 = 10.40.

Discussion sur P1 (note 0) : le barème propose P1 = 15mn42. Cela correspond bien au niveau des candidates les plus faibles : 18 candidates sur 837 (abandons compris) ont fait moins bien. Offrir une cote basse accessible à tous est une condition de motivation pour que les candidates les plus faibles se préparent positivement. Nous retiendrons P3 = 16.00.

Remarques :

En annexe, on trouvera une comparaison de ces résultats avec les indications fournies par les tables du test de Cooper (sur 12 minutes) d'une part, et avec les résultats de lycéens et lycéennes en situation d'évaluation. On verra en se reportant à ces annexes que les barèmes déduits des résultats des candidats convergent assez bien avec ces informations. Il faut toutefois remarquer que les candidats ont généralement réalisé de meilleures performances que celles qu'on pouvait attendre à partir des estimations liées au Cooper ou aux évaluations lycéennes. Cela s'explique sans doute pour une bonne part par la motivation liée à un concours.

3.3. Barème proposé

NOTES	PERF
20.0	8.00
19.0	8.11
18.0	8.24
17.0	8.37
16.0	8.51
15.0	9.06
14.0	9.22
13.0	9.40
12.0	9.59
11.0	10.19
10.0	10.40

NOTES	PERF
9.0	11.03
8.0	11.28
7.0	11.54
6.0	12.22
5.0	12.53
4.0	13.25
3.0	14.00
2.0	14.37
1.0	15.18
0.0	16.00

Cet ajustement du barème modifie peu le bilan des notes (moyenne = 10.05, écart type = 3.46, asymétrie = -0.008).

3.4. Évolutions possibles, variantes

a) Le barème sous-estime sans doute le niveau des meilleures candidates en raison des conditions de déroulement de l'épreuve. Il faut prévoir une évolution vers le haut.

b) Si l'on veut obtenir une dispersion des notes plus grande (écart type des notes = 4), sans changer la moyenne, nous obtenons :

NOTES	0.0	5.0	10.0	15.0	20.0
PERF	15.02	12.31	10.39	9.16	8.15

Pour la majorité des candidats, la dispersion des notes est plus grande mais au niveau 0 et 20 le nombre de candidats non différenciés est plus important. Il est plus facile de gagner un point (écart de performance moindre) ce qui augmente la fonction incitative du barème pour la majorité des candidats, mais au niveau des notes 0 et 20 cette fonction incitative est diminuée. Malgré les avantages de cette formule, nous ne la proposons pas en raison de ses inconvénients au niveau des cotations extrêmes.

c) Le niveau des meilleures candidates étant vraisemblablement sous-estimé, il serait donc intéressant de prévoir une notation sur 16 + 5 points qui permettrait aux meilleurs candidats de pouvoir exprimer leur niveau réel. On aurait alors en transposant la notation sur 20 points sur une échelle de 0 à 15 et en prolongeant cette notation d'un point la solution suivante (la note 7.5 non présente sur ce résumé correspond à 10.40).

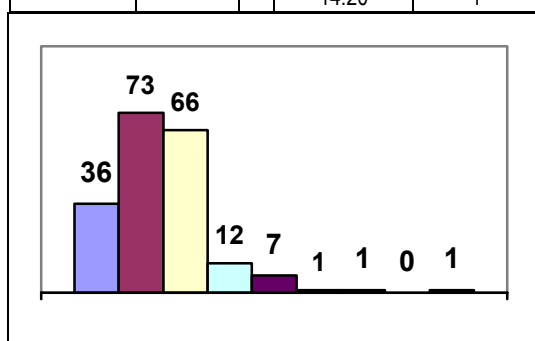
NOTES	PERF
16	7.46
15	8.00
14	8.15
13	8.32
12	8.51
11	9.12
10	9.34
9	9.59
8	10.26

NOTES	PERF
7	10.55
6	11.28
5	12.03
4	12.42
3	13.25
2	14.12
1	15.04
0	16.00

4. CRPE hommes : les temps sur 2000 m

4.1. Étude des performances

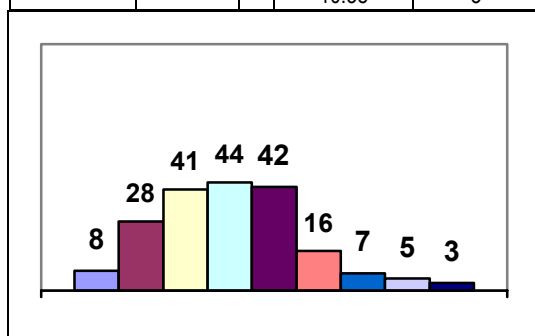
Résumé A		Lim. Sup.	Effectifs
Effectif	197	7.16	36
Moyenne	8.078	8.09	73
Médiane	8.020	9.02	66
Ec type	0.591	9.55	12
Asymétrie	1.909	10.48	7
Minimum	6.23	11.41	1
Maximum	14.20	12.34	1
		13.27	0
		14.20	1



L'effectif est assez important (197 + 1 abandon) ce qui permet d'avoir une idée assez précise du niveau des candidats. Les performances sont très dispersées, de 6mn23s à 14mn20s (un seul abandon) mais au-dessus de 10mn48s on n'a que trois performances particulièrement faibles.

L'asymétrie des performances est forte (histogramme, coefficient d'asymétrie, médiane différente de moyenne). Mais il faut toutefois que les 3 performances très faibles influencent fortement l'asymétrie. Si on défalque ces trois performances, l'asymétrie est confirmée mais moindre.

Résumé B		Lim. Sup.	Effectifs
Effectif	194	6.51	8
Moyenne	8.037	7.20	28
Médiane	8.020	7.48	41
Ec type	0.485	8.16	44
Asymétrie	0.512	8.45	42
Minimum	6.23	9.13	16
Maximum	10.38	9.41	7
		10.10	5
		10.38	3



Nous rechercherons un barème sur la base ces 194 performances afin de limiter l'influence de ces trois résultats très faibles.

4.2. Recherche du barème et conséquences sur les notes

Le barème proposé pour obtenir une normalisation des 194 notes peut être résumé ainsi :

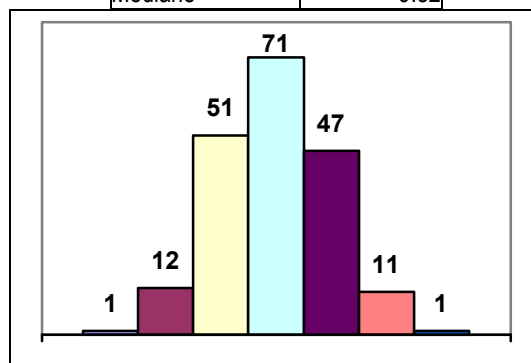
NOTES	0	5	10	15	20
PERF	11.07	9.21	7.59	6.57	6.09
	P3		P2		P1

Remarque : Si l'on calcule avec les 197 données on obtient le résumé suivant. L'influence des 3 performances très faibles est donc forte notamment sur les cotations basses.

NOTES	0	5	10	15	20
PERF	12.32	9.41	7.58	6.58	6.21
	P3		P2		P1

On obtient alors la distribution suivante des 194 notes.

Moyenne	10.00
Ec type	3.33
Asymétrie	-0.010
Mini =	1.3
Maxi =	18.4
Médiane	9.82



(méthode de répartition en classes expliquée plus haut en 3.2.)

Discussion sur P3 (note 20) : On remarque que la meilleure note est 18.4 pour 6mn23s (équivalence 1000 m selon Letessier 1992, 2mn49s). Le barème actuel 2003 (assez facile pour les garçons) indique 6mn50s (équivalence 1000 m : 3mn06s). Il semble prudent, pour ne pas introduire de rupture trop importante avec les barèmes précédents, de retenir 6mn30s (soit 2mn57s au 1000 m).

Discussion sur P2 (note 10) : le barème propose 7mn59s (équivalence 1000 m : 3.35). Cette proposition paraît bien en accord avec le niveau médian des candidats, le barème actuel 2003 indique 8.40 (pour la note 7.5) ce qui paraît assez facile (3.51 au 1000 m). Nous retiendrons P2 = 8.00.

Discussion sur P1 (note 0) : le barème propose P1 = 11.07. Cela semble bien correspondre au niveau des candidats les

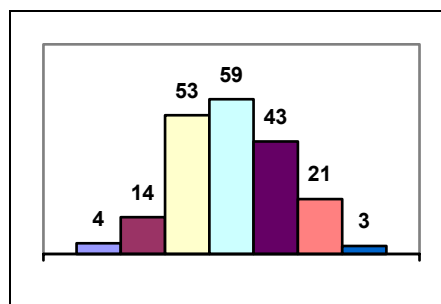
plus faibles. Le barème CRPE 2003 indique 11mn pour la note 0. Nous retiendrons P3 = 11.00

4.3. Barème proposé et distribution des notes

NOTES	PERF
20.0	6.30
19.0	6.36
18.0	6.43
17.0	6.51
16.0	6.59
15.0	7.07
14.0	7.16
13.0	7.26
12.0	7.37
11.0	7.48
10.0	8.00

NOTES	PERF
9.0	8.13
8.0	8.27
7.0	8.42
6.0	8.58
5.0	9.15
4.0	9.33
3.0	9.53
2.0	10.14
1.0	10.36
0.0	11.00

Moyenne	10.01
Ec type	4.17
Asymétrie	-0.213
Minimum	-6.4
Maximum	21.2
Médiane	9.84
Notes >= 20	1.5%
Notes >= 10	46.2%
Notes < 0	1.0%



On voit que les ajustements opérés sur P1, P2 et P3 modifient peu les médianes et moyennes mais elles augmentent la dispersion ce qui nous paraît positif pour un concours. Nous faisons l'hypothèse que ce barème proposant des objectifs plus élevés au niveau des moyens et des meilleurs provoquera une élévation des performances pour les meilleurs candidats qui étaient freinés par le barème 2003.

4.4. Évolutions possibles, variantes

- Le barème sous-estime sans doute le niveau des meilleures candidates en raison des conditions de déroulement de l'épreuve. Il faut prévoir une évolution vers le haut.
- Comme pour les filles, le niveau des meilleurs candidats est vraisemblablement sous-estimé, il est donc intéressant de prévoir une notation sur 16 + 5 points.

5. CRPE femmes : les écarts au projet

Le barème prévoit 0 point lorsque l'écart au projet est supérieur à 30 secondes. Dans ces conditions 34% des candidates obtiennent la note 0. Les raisons sont multiples :

- Le barème est mal adapté
- Le règlement de l'épreuve ne permet pas aux candidates faibles (hors barème performance) d'obtenir une note non nulle à l'écart au projet. Si le barème « performance » est trop difficile pour les candidates faibles, celles-ci sont aussi pénalisées à la note d'écart au projet.

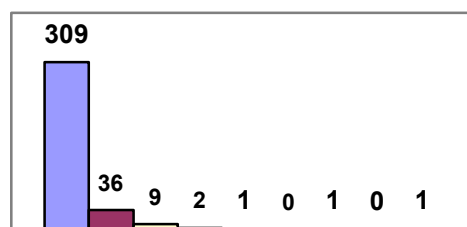
Il faut donc raisonner dans le cadre de la réglementation actuelle mais surtout dans le cadre d'une réglementation améliorée :

- par un barème performance mieux adapté à l'ensemble des candidates
- par la possibilité de se fixer un temps hors barème performance pour les meilleures comme pour les plus faibles. Dans ces conditions les écarts au projet pourront être mieux adaptés ce qui permet d'ajuster le niveau d'exigence.

5.1. L'étude des performances

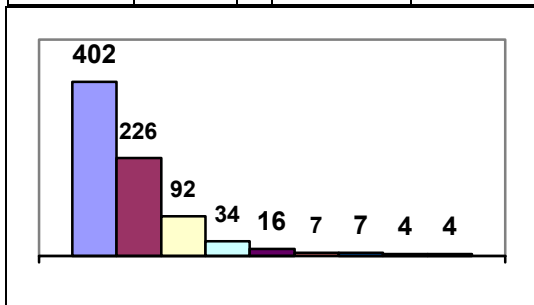
L'histogramme montre bien que 11 écarts sont nettement supérieurs aux autres.

Résumé		Lim. Sup.	Effectifs
Effectif	825	53.3	736
Moyenne	26.584	106.7	64
Médiane	17.000	160.0	17
Ec type	37.744	213.3	3
Asymétrie	6.138	266.7	2
Minimum	0	320.0	0
Maximum	480	373.3	1
		426.7	0
		480.0	2



Compte tenu des conditions de l'épreuve, nous n'avons retenu que les écarts des candidates n'ayant pas automatiquement 0 à l'épreuve de 2000 m parce qu'elles réalisaient plus 13mn30. Restent 792 candidates.

Résumé		Lim. Sup.	Effectifs
Effectif	792	16.0	402
Moyenne	22.164	32.0	226
Médiane	16.000	48.0	92
Ec type	21.815	64.0	34
Asymétrie	2.177	80.0	16
Minimum	0.00	96.0	7
Maximum	144.00	112.0	7
		128.0	4
		144.0	4



On voit que la distribution est extrêmement asymétrique (comme dans toutes les tâches de ce type). Alors que la mesure est bornée d'un côté (le meilleur score est nécessairement 0), on enregistre 23 écarts nuls, 21 écarts de 1 seconde, 28 écarts de 2 secondes. Face à de tels résultats on peut s'interroger sur les conditions de la mesure : ces candidates ont-elles été mieux renseignées que les autres durant la course ? Il semble que certains jurys se sont adaptés face à cette situation très contestable.

On ne voit pas en effet comment sans aucune information 72 candidates (soit 8.6% des 837 candidates) ont pu obtenir un écart inférieur ou égal à 2 secondes sur un 2000 mètres. Des athlètes ayant des années d'expérience ne parviennent pas à un tel résultat sans temps de passage.

Un tel échantillon pose donc de sérieux problèmes pour en déduire un barème.

5.2. Recherche d'un barème

Il faut se placer dans deux perspectives : celle où la réglementation ne change pas au plan de l'information sur les temps de passage (il faut alors considérablement modifier le barème) ; celle où la réglementation est améliorée et il faut modifier le barème dans une moindre mesure.

a) Sans communication des temps de passage

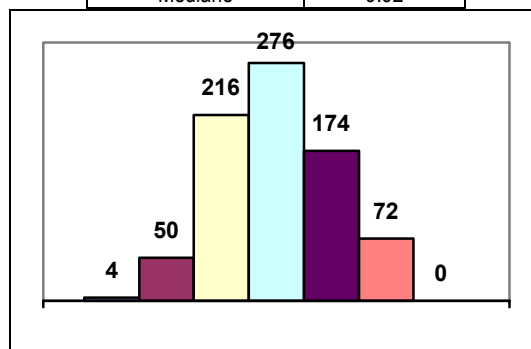
On suppose toutefois que les barèmes concernant la performance (temps à l'arrivée) sont néanmoins revus dans le sens proposé plus haut (voir 3.3.).

Le barème pour que le coefficient d'asymétrie soit pratiquement nul est le suivant.

NOTES	0	5	10	15	20
PERF	190.3	58.0	15.6	2.1	-2.2

On a alors la distribution des notes qui suit.

Moyenne	10.00
Ec type	3.33
Asymétrie	-0.007
Mini =	1.2
Maxi =	16.8
Médiane	9.92

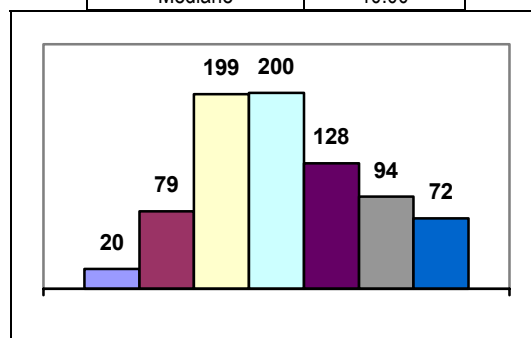


Cette solution n'est pas satisfaisante (performance négative pour la note 20 ce qui est impossible), il faut donc l'adapter. L'une des solutions possibles consiste à adopter P1=120 (soit 2 minutes d'écart au projet), P2=16 (c'est la médiane des performances), P3=2 (pour avoir 20, l'écart ne doit pas être supérieur à 2 secondes, ce qui peut paraître exigeant mais 72 candidates ont des écarts inférieurs ou égaux à 2 s).

On obtient alors le barème et le bilan des notes suivants.

NOTES	0	5	10	15	20
PERF	120.0	43.9	16.0	5.8	2.0

Moyenne	11.03
Ec type	6.17
Asymétrie	1.265
Mini =	-0.9
Maxi =	32.5
Médiane	10.00



Cette solution est un compromis entre les données empiriques, issues d'un échantillon qui pose problème compte tenu des règles de l'épreuve, et la réflexion sur ce qui devrait se passer si le barème de la performance est adapté.

Mais ce qui est en cause plus fondamentalement comme nous l'avons vu, ce sont les conditions d'information durant l'épreuve. Si ces conditions changent, les résultats observés dans d'autres conditions ne sont plus aussi pertinents.

b) Avec communication des temps de passage

Le barème de performances évolue dans le sens indiqué plus haut (voir 3.3.) et les temps de passage sont communiqués aux candidats tous les 400 m.

P1, P2 et P3 sont alors estimés en fonction de l'expérience de cette année et d'hypothèses vraisemblables. Le barème de novembre 2002 paraît une bonne base de départ si des temps de passage sont fournis tous les 400 mètres. Mais 72 candidates ayant eu des écarts inférieurs ou égaux à 2, il est justifié de rendre plus difficile la note 20 que dans le barème 2002-2003. Le niveau médian observé peut être conservé, il est d'ailleurs proche des indications du barème 2003, et le niveau inférieur peut être légèrement assoupli.

Nous aurions alors :

NOTES	0	5	10	15	20
PERF	40.0	26.4	16.0	8.1	2.0

Sur la base des résultats de cette année (les 792 candidates non condamnées à une note nulle), la moyenne serait alors de 9 avec un écart type très important (8.4). Mais avec des informations tous les 400 m et des barèmes de performances mieux adaptés, les candidates auront des projets mieux ajustés et pourront mieux les maîtriser. Les écarts au projet seront alors beaucoup plus petits ce qui aura pour conséquence d'élever la moyenne des notes et d'en réduire l'écart type. On peut penser que dans ces conditions, la moyenne des notes sera voisine de 10 et que l'écart type des notes reviendra à une valeur comprise entre 3 et 4.5.

Les règles ayant été adaptées dans le sens souhaité, une nouvelle étude pourra être faite sur les nouveaux écarts au projet et il sera alors possible de stabiliser le barème en fonction des nouvelles observations.

Nous donnerons plus loin la conversion de ce barème selon une échelle de 0 à 5 points mais soulignons dès maintenant qu'il est très souhaitable que la notation soit prévue au demi-point près, comme cela est prévu pour la performance.

5.3. Relation entre la performance et l'écart au projet

Si l'on calcule la corrélation entre le temps réalisé et l'écart au projet, on obtient $r=0.56$ pour les 825 candidates. Apparemment, plus on court lentement et plus l'écart au projet est grand. Mais les distributions ne sont pas gaussiennes et le nuage de points montre que la relation n'est pas simple.

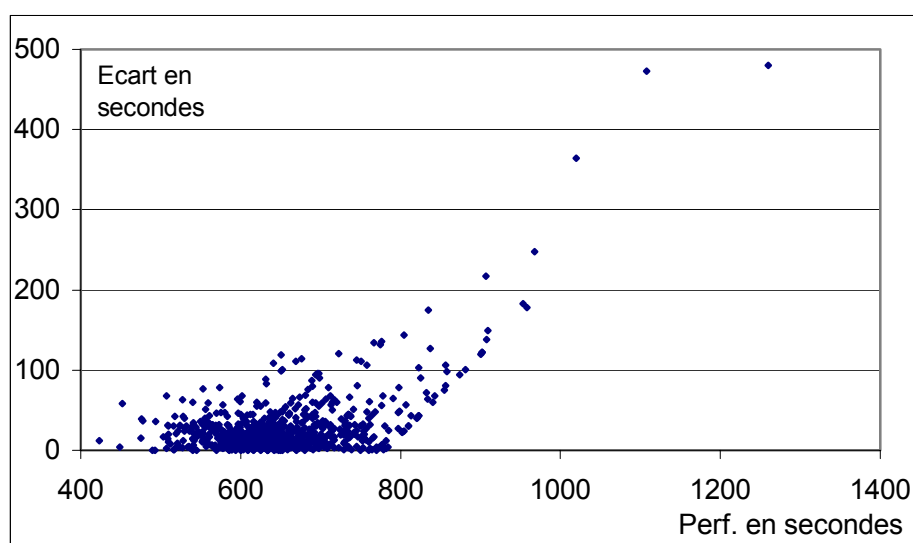
Si l'on divise les 825 candidates en quatre groupes (de 206 ou 207 sujets) en fonction de leur temps réalisé à l'arrivée du 2000 mètres, on obtient les résultats du tableau suivant.

Niveaux (temps 2000)	Valeur de r	Corrélation	Commentaires
Supérieur	-0.16	faible	Plus on court vite et plus l'écart est grand
Moyen sup.	0.09	nulle	Pas de relation
Moyen inf.	0.05	nulle	Pas de relation
Inférieur	0.79	forte	Plus on court lentement et plus l'écart est grand
TOUTES	0.56	moyenne	Plus on court lentement et plus l'écart est grand

Pour les trois quarts des candidats au moins, la relation tend à être nulle entre la performance et l'écart au projet. En revanche pour le quart inférieur, cette relation est forte à très forte (0.79). On observe même $r = 0.90$ pour les 10% des performances les plus faibles).

Pour les meilleurs on pouvait s'attendre à une corrélation positive : il n'en est rien, la corrélation faiblement négative est statistiquement significative en raison de l'effectif élevé (206). Deux explications peuvent être avancées : à ce niveau, un plus grand nombre de candidates prennent le risque de s'écarter de leur projet en recherchant la meilleure performance ; de plus le règlement actuel a pénalisé quelques candidates hors barème vers le haut.

En résumé, la corrélation observée (0.56) est liée à une très forte corrélation pour 10% des candidates les plus faibles. Pour la majorité des candidates la corrélation entre performance et écart au projet est nulle, ce qui montre que les deux évaluations ne sont pas redondantes.



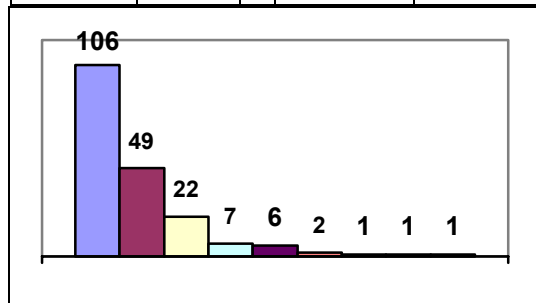
6. CRPE hommes : les écarts au projet

Il faut d'abord observer que seuls deux temps sont supérieurs à 11.30 et correspondent à des notes globales égales à 0 en fonction des règles indiquées plus haut (voir 2.). Par ailleurs on note un seul abandon. Alors que 6.1% des candidates ont eu comme note totale 0, ce pourcentage est de 1.5% pour les garçons. On peut donc dire que les règles de l'épreuve sont mieux adaptées aux garçons qu'aux filles, parce que le niveau de performances a été mieux estimé pour les garçons que pour les filles.

Nous allons voir que cela a des conséquences sur les écarts au projet. Nous ne ferons l'analyse que sur les 195 candidats dont les temps sont inférieurs à 11.30 et qui pouvait donc se fixer un projet en rapport avec leurs possibilités.

6.1. L'étude des écarts au projet

Résumé		Lim. Sup.	Effectifs
Effectif	195	10.7	106
Moyenne	14.031	21.3	49
Médiane	10.000	32.0	22
Ec type	14.498	42.7	7
Asymétrie =	2.374	53.3	6
mini	0.00	64.0	2
maxi	96.00	74.7	1
		85.3	1
		96.0	1



Comme pour les candidates la distribution est fortement asymétrique mais alors que la médiane était de 16 pour les filles elle n'est que de 10 pour les garçons. Par ailleurs on relève 31 écarts inférieurs ou égaux à 2 secondes (soit 15.6% des candidats alors que ce pourcentage était de 8.6% pour les filles).⁷

On peut donc dire que les garçons ont des scores nettement supérieurs à ceux des filles dans l'écart au projet ce qui ne justifie pas empiriquement la solution de barème mixte qui a été adoptée.

Nous reviendrons plus loin sur ce point notamment en exprimant l'écart en valeur relative par rapport au temps réalisé mais on peut remarquer déjà que si l'on réalise les rapports médiane écart au projet / médiane temps à l'arrivée, on obtient

⁷ On remarque que 53 de l'ensemble des 198 candidats hommes (27%) ont des écarts inférieurs ou égaux à 5 secondes ce qui leur donne la note maximale. Ce niveau est exceptionnel... si les prestations ont été réalisées sans aucune information durant la course.

en pourcentage un écart relatif de 2.1 pour les garçons et 2.5 pour les filles.

6.2. Recherche d'un barème

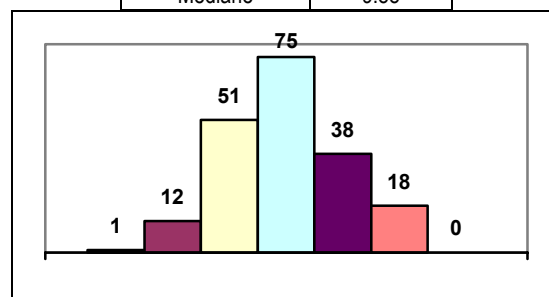
Comme pour les filles, il faut distinguer deux cas : lorsque les temps de passage ne sont pas donnés aux candidats et lorsqu'ils sont donnés.

a) Sans temps de passage.

La recherche de Po tel que l'asymétrie des notes soient pratiquement nulle conduit au barème (résumé) et bilan des notes suivants

Notes	0	5	10	15	20
Perf	127.4	37.2	9.7	1.2	-1.4

Moyenne	10.00
Ec type	3.33
Asymétrie	-0.008
Mini =	1.2
Maxi =	16.7
Médiane	9.88

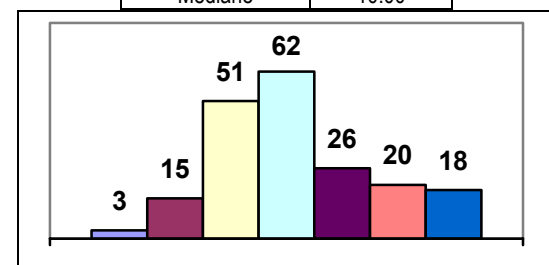


Les conditions de la mesure et le nombre très important d'écarts nuls expliquent qu'une cotation négative soit proposée pour le niveau le plus haut. Il faut donc adapter la proposition de barème.

Si comme pour les filles on attribue à P2 la valeur de la médiane, on parvient à une solution comparable aux filles avec P1 = 90 et P3 = 1.

Notes	0	5	10	15	20
Perf	90.0	30.1	10.0	3.3	1.0

Moyenne	11.05
Ec type	5.92
Asymétrie	1.301
Minimum	-0.3
Maximum	29.6
Médiane	10.00



Mais comme ce déroulement sans temps de passage est très discuté, nous proposons une solution avec temps de passage.

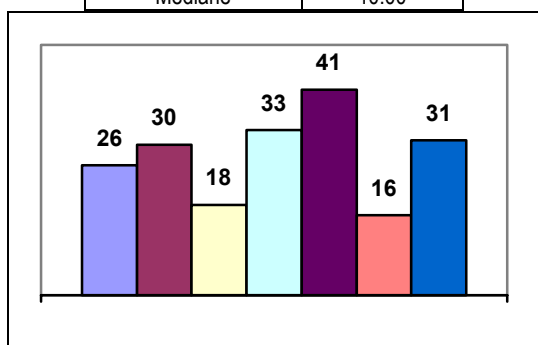
b) Avec temps de passage

Selon la même logique que pour les filles, mais en tenant que les garçons réalisent des écarts nettement inférieurs à ceux des filles, nous proposons : P1 = 30 (note 0), P2 = 10 (note 10) ; P3 = 1.5 (note 20).

Nous avons alors le barème et le bilan des notes suivants.

NOTES	0	5	10	15	20
PERF	30.0	17.9	10.0	4.9	1.5

Moyenne	9.78
Ec type	7.59
Asymétrie	-0.314
Minimum	-12.4
Maximum	23.2
Médiane	10.00



On pourra s'étonner de constater que la distribution est si éloignée d'une distribution gaussienne puisque c'est notre choix initial. Mais il faut rappeler que la simulation est faite avec des données réalisées dans de toutes autres conditions. Si les candidats sont informés tour par tour, les plus mauvais écarts

seront améliorés et les notes convergeront vers le centre de la distribution. En revanche on devrait continuer à observer un grand nombre de bonnes notes. La solution serait d'adopter 0 pour la valeur P3. Nous ne le faisons pas parce que nous avons déjà considérablement « durci » les exigences précédentes (écart de 5 s ou moins).

6.3. Relation entre la performance et l'écart au projet

Si l'on calcule la corrélation entre le temps réalisé et l'écart au projet, on obtient $r=0.51$ pour les 197 candidats. Plus on court lentement et plus l'écart au projet est grand. Comme pour les femmes ($r = 0.56$).

Et le nuage points est de même forme que celui des femmes. Nous avons donc calculé les corrélations par bloc d'un quart des données

Niveaux (temps 2000)	Valeur de r	Relation	Commentaires
Supérieur	-0.19	faible	Plus on court vite et plus l'écart est grand
Moyen sup.	0.02	nulle	Pas de relation
Moyen inf.	0.37	assez faible	Plus on court lentement et plus l'écart est grand
Inférieur	0.77	forte	
TOUTES	0.51	moyenne	Plus on court lentement et plus l'écart est grand

Sauf pour la classe « moyen inférieur », les corrélations sont très proches de celles des filles. Les commentaires faits pour les filles valent donc pour les garçons.

7. Les écarts au projet : barème mixte ?

L'étude des écarts au projet montre que les filles ont des écarts plus grands que les garçons.

	Filles (N=825)	Filles (N=792)	Garçons (N=195)
Moyenne	26.58	22.16	14.03
Médiane	17	16	10
Minimum	0	0	0
Maximum	480	144	96

Ces différences correspondent sensiblement à un écart d'environ 1.5 point en notation sur 5 selon les normes actuelles (soit une différence de près de 6 points en notation sur 20 !).

Cette étude montre qu'on devrait proposer un barème pour les filles et un barème pour les garçons, car sans information durant la course les erreurs ont tendance à être d'autant plus grandes que le temps de course est plus grand.

Toutefois on peut penser que ces différences seront considérablement atténuées avec une information tous les 400 mètres. On ne peut donc se fonder pleinement sur les observations actuelles réalisées dans des conditions qu'il faudra améliorer.

Il faut aussi remarquer que les différences sont partiellement atténuées lorsqu'on rapporte l'écart au projet au temps à l'arrivée. L'étude sur ces écarts relatifs ne permet cependant pas de parvenir à des barèmes mixtes.

Nous parvenons donc à la conclusion suivante :

- Si l'épreuve reste telle qu'elle est au plan de la communication des temps de passage, les écarts au projet sont fortement influencés par le temps de course et des barèmes par sexe s'imposent (voir plus haut 6.2.a et 5.2.a).
- Si l'épreuve est améliorée au niveau des barèmes de performances et de l'information sur les temps de

passage, nous faisons l'hypothèse que les écarts des filles se rapprocheront sensiblement de ceux des garçons. Les filles cesseront d'être nettement désavantagées par un barème commun. On peut avoir dans ce cas le barème suivant, intermédiaire entre les solutions préconisées plus haut pour les filles et les garçons.

Notes	0	5	10	15	20
Perf	35	21.1	12	6	2

Rappel : en juin 2003, 16% des garçons et 9% des filles ont eu des écarts au projets égaux ou inférieurs à 2 secondes (sans information, en principe, sur leur temps de passage). Fixer la note 20 pour un écart inférieur ou égal à 2 secondes est donc accessible et cela correspond à un très bon niveau de maîtrise des allures en demi-fond.

Conclusions et résumé des propositions

Bien que certains formateurs préfèrent une épreuve sur 3000 mètres ou une course de longue durée, le 2000 mètres au CRPE est une épreuve intéressante si les dispositions concernant le déroulement des épreuves sont améliorées à la lumière des observations des enseignants et de l'étude des performances réalisées par plus de 800 filles et près de 200 garçons en situation de concours. Nous ramasserons nos conclusions à travers une proposition de réglementation puis de quelques commentaires.

Propositions

1. Les candidates et les candidats réalisent un 2000 mètres sur une piste de 400 mètres. La durée de l'épreuve est limitée à 20 minutes pour les femmes et à 15 minutes pour les hommes.
2. Ils courent sans chronomètre ni montre personnelle mais sont informés par le jury de leurs temps de passage tous les 400 mètres.
3. Le temps à l'arrivée de chaque candidat est enregistré au dixième de seconde près.
4. Avant le départ, chaque candidat(e) communique au jury le temps qu'il projette de réaliser. Ce temps sera comparé au temps réalisé et la différence qu'elle soit positive ou négative constituera son écart au projet. Un candidat peut avoir un projet en dehors des limites du barème. Exemple : une candidate vise 7mn30s, elle réalise 7mn32, son écart au projet est 2 secondes.
5. Les candidats sont alors évalués en fonction de leur performance (temps à l'arrivée) sur 16 points, et de leur écart au projet sur 5 points. Le total des points est donc sur 21 points mais la note finale est ramenée à 20 si ce total dépasse 20. Cette disposition permet de mieux tenir compte de l'excellence des niveaux en performance ou en écart au projet.
6. Pour la performance les barèmes sont différenciés selon le sexe. Pour l'écart au projet, le barème est mixte. Les deux évaluations sont réalisées au demi-point près et arrondies à la cotation égale ou inférieure. Exemple pour des candidates : 8mn03 donne la note 14.5 ; 11mn11 donne la note 6.5.

Pourquoi

Cette réglementation de l'épreuve s'appuie sur l'étude de 837 résultats de candidates et de 198 hommes en situation de

concours. Elle intègre les remarques des enseignants souhaitant notamment :

- la disparition de la disposition obligeant un candidat à adopter un projet dans les limites du barème ;
- la possibilité de donner des temps de passage à chaque 400 m ;
- la valorisation de la performance sans minorer le poids de la maîtrise du projet.

Pour parvenir à cette proposition nous avons d'abord calculé des barèmes sur 20 points avec les conventions et considérations suivantes :

- Dans un concours, il importe de différencier les candidats. Ce n'est donc pas la moyenne des notes qui compte le plus mais la dispersion des notes. Il faut aussi pouvoir différencier les candidats dans les extrémités ce qui impose de rendre exceptionnelles les notes 20 et 0.
- Nous avons alors établi des barèmes sur 20 points en tendant vers une distribution gaussienne des notes (moyenne = 10 et écart type = 3.33). Ces choix permettent de situer les notes 20 et 0 à 3 écarts types de la moyenne et donc de prendre en compte 99.7% des cas. Trois cas sur 1000 sont donc théoriquement « hors barème » et ramenés à 20 et 0. Il faut toutefois remarquer que nous n'avons pas situé les notes 20 au niveau où il aurait été juste de les situer dans cette perspective, car cela aurait créé une trop grande rupture avec les dispositions antérieures et les représentations majoritaires des formateurs.
- Selon cette démarche, les niveaux d'exigence pour l'écart au projet et pour la performance sont comparables, en fonction de ce qui a été observé en juin 2003.
- Puis nous avons transposé ces barèmes dans les échelles de 15 points pour la performance et de 5 points pour l'écart au projet.
- Enfin nous avons amélioré la différenciation aux extrémités en créant un bonus performance pour les meilleurs (notation sur 16 points) et une pénalisation en cas de performance très faible. Alors qu'en juin 2003 une candidate obtenait nécessairement 0 si elle dépassait 13 minutes 30, une candidate doit maintenant dépasser 20 minutes pour avoir 0. Cette disposition permet aussi de borner en temps l'épreuve ce qui ne serait pas possible avec des projets totalement libres.

- Notons enfin que l'étude des écarts au projet montre qu'il serait souhaitable d'adopter des barèmes par sexe. Nous ne l'avons pas fait parce que les modifications de déroulement proposées ci-dessus devraient atténuer les handicaps des candidates et rapprocher leurs résultats de ceux des candidats hommes.

Remarque

Si l'on intègre les abandons, avec les barèmes ci-dessous le bilan des notes de performances (sur 20) est le suivant :

- $m = 9.94$ et $\sigma = 4.10$ pour les hommes,
- $m = 9.92$ et $\sigma = 3.54$ pour les femmes.

Ces moyennes sont inférieures de 1 à 2 points à celles de danse et badminton dans trois départements. L'écart type est comparable pour les hommes mais un peu plus faible pour les femmes. On obtient pour les femmes un écart type des notes de 4.04 si P1 = 8.20, P = 10.40 et P3 = 15.00.

Barème 2000 m Femmes			
NOTES	PERF	NOTES	PERF
16.0	7.46	7.5	10.40
15.5	7.53	7.0	10.55
15.0	8.00	6.5	11.11
14.5	8.08	6.0	11.28
14.0	8.15	5.5	11.45
13.5	8.24	5.0	12.03
13.0	8.32	4.5	12.22
12.5	8.42	4.0	12.42
12.0	8.51	3.5	13.03
11.5	9.01	3.0	13.25
11.0	9.12	2.5	13.48
10.5	9.23	2.0	14.12
10.0	9.34	1.5	14.37
9.5	9.46	1.0	15.04
9.0	9.59	0.5	15.31
8.5	10.12	0.0	16.00
8.0	10.26		

Barème 2000 m Hommes			
NOTES	PERF	NOTES	PERF
16.0	6.22	7.5	8.00
15.5	6.26	7.0	8.09
15.0	6.30	6.5	8.17
14.5	6.34	6.0	8.27
14.0	6.39	5.5	8.37
13.5	6.43	5.0	8.47
13.0	6.48	4.5	8.58
12.5	6.53	4.0	9.09
12.0	6.59	3.5	9.21
11.5	7.04	3.0	9.33
11.0	7.10	2.5	9.46
10.5	7.16	2.0	9.59
10.0	7.23	1.5	10.13
9.5	7.30	1.0	10.28
9.0	7.37	0.5	10.44
8.5	7.44	0.0	11.00
8.0	7.52		

Barème écart au projet (mixte)		Pénalisations FEMMES au-delà de 16 minutes	Pénalisations HOMMES au-delà de 11 minutes
NOTES	PERF		
5.0	2.0	0.5 pt : de 16 à 16.29 compris	0.5 pt : de 11 à 11.29 compris
4.5	3.4	1.0 pt : de 16.30 à 16.59 compris	1.0 pt : de 11.30 à 11.59 compris
4.0	5.0	1.5 pt : de 17 à 17.29 compris.	1.5 pt : de 12 à 12.29 compris.
3.5	7.0	2.0 pts : de 17.30 à 17.59 compris.	2.0 pts : de 12.30 à 12.59 compris.
3.0	9.3	2.5 pts : de 18 à 18.29 compris.	2.5 pts : de 13 à 13.29 compris.
2.5	12.0	3.0 pts : de 18.30 à 18.59 compris.	3.0 pts : de 13.30 à 13.59 compris.
2.0	15.2	3.5 pts : de 19 à 19.29 compris.	3.5 pts : de 14 à 14.29 compris.
1.5	19.0	4.0 pts : de 19.30 à 20.00 compris.	4.0 pts : de 15.30 à 16.00 compris.
1.0	23.5	Au-delà de 20 minutes, pénalisation de 5 pts	Au-delà de 16 minutes, pénalisation de 5 pts
0.5	28.8		
0.0	35.0		

Exemples pour 3 candidates (A, B, C) qui dans la réglementation 2003 obtenaient toutes 0/20.
A prévoit 14.12, elle réalise 14.08. Écart au projet = 5 s. Note perf = 2.0, note projet = 4. Total = 6.
B prévoit 16.45, elle réalise 16.50. Écart au projet = 5 s. Note perf = 0, note projet = 4-1. Total = 3.
C prévoit 19.40, elle réalise 19.38. Écart au projet = 2 s. Note perf = 0, note projet = 5-4. Total = 1.

Bibliographie sommaire

Cleuziou JP (1996), *Mesure de l'asymétrie et cotation des performances*, Revue Science et motricité n° 28, Paris. (exposé de la méthode de calcul des barèmes en annulant l'asymétrie des notes).

Goetghebuer G et Cazorla G (2003), Tables par centilage du test de Cooper, Revue Sport et Vie n° 76.

Annexe 1 : comparaison avec d'autres résultats

Comparaison avec les tables du test de Cooper

Le test de Cooper consiste comme on le sait à couvrir la plus grande distance possible en 12 minutes. Depuis plus de 30 ans, on estime que des centaines de millions de personnes (principalement adultes) ont réalisé ce test. L'Institut pour la recherche en aérobic de Dallas a mémorisé ces résultats dans de très importantes bases de données, ce qui permet de déboucher sur des tables de niveau des individus par centilage. La revue Sport et Vie (n° 76, janvier-février 2003) a publié ces tables par centiles (ou percentiles). Nous ne retiendrons que trois cotes qui permet d'estimer le niveau médian (C50), le niveau supérieur (C99) et le niveau très faible (C01) pour les femmes et les hommes de 20 à 29 ans.

Notons que C01 correspond sensiblement à la note 1.5 dans une notation sur 15 points ($m = 7.5$ et $\sigma = 2.5$) : la note z de 1.5 ($(1.5 - 7.5) / 2.5 = -2.5$) correspond au centile C00.8 ce qui est très proche de C01. De même C99 correspond sensiblement à 13.5. (0 correspond au centile C00.13 et 20 au centile C99.87 dans ce système de notation)

Centiles	% de performances inférieures	Femmes	Hommes
C99	99%	2864	3121
C50	50%	2076	2414
C01	1%	1512	1706

Ces tables concernent donc 98% des sujets.

Il faut remarquer que les sujets qui ont réalisé ce test n'étaient pas placés dans les mêmes conditions de motivation que des candidats à un concours ; on ne peut donc retenir sans précaution ces indications toutefois utiles.

Méthode.

Le calcul des équivalences entre distances courues pendant 12 minutes et temps réalisés sur 2000 m n'est pas simple car les distances courues lors d'un test de Cooper varient très fortement : ici de 1512 m à de 3121 m.

On peut estimer ces équivalences à partir d'équivalences données par des tables de cotation sur 1500, 2000 et 3000 m. Ainsi on observe que temps sur 2000 m et temps sur 1500 m sont dans un rapport R pratiquement constant (1.4, voir plus loin) aux niveaux de pratique concernés ici. Connaissant le temps sur 1500 m on peut donc en déduire la valeur sur 2000 m. Si un sujet court 1500 m en 12 minutes et si $R = 1.4$ (comme nous le verrons plus loin), le temps équivalent sur 2000 m est estimée à 16.48.

Mais la distance courue sur 12 minutes n'est pas stable, il faut donc calculer des équations permettant d'estimer R en fonction de la distance courue : 1650 m, 1740 m ... Cela est possible à partir de quelques équivalences que nous pouvons trouver par exemple dans une table comme la table Letessier. Nous débouchons alors sur deux équations.

Équation A (1500 à 2000 m). Lorsque la distance courue est 1500 m, on obtient une estimation convenable du temps sur 2000 m en multipliant le temps du 1500 m (en secondes) par un coefficient R égal à 1.40 (1.40 = moyenne des rapports entre temps sur 2000 et temps équivalents sur 1500 de la table Letessier).

Mais plus la distance (D) courue se rapproche de 2000 m et plus ce coefficient se rapproche de 1. On dispose donc de deux couples (1500, 1.40) et (2000, 1.00), ce qui permet de préciser l'équation du coefficient R : $R = -0.8 D + 2.6$ lorsqu'on exprime D en km et non en mètres. On multiplie le temps réalisé sur la distance D par R pour obtenir l'équivalence en temps sur 2000.

Comme le temps couru dans le test de Cooper est toujours de 12 minutes (720 s) il suffit de multiplier C par 720. Autrement dit : $T_{2000} = 720 \times (-0.8 D + 2.6)$, avec T_{2000} en secondes et D en kilomètres.

Équation B (2000 à 3000 m). On montre de la même manière qu'on peut estimer T_{2000} en multipliant le temps sur 3000 m par $R = 0.634$. À partir de là, on précise une deuxième équation : $T_{2000} = 720 \times (-0.366 D + 1.732)$ avec T_{2000} en secondes et D en kilomètres.

Ces deux équations permettent de calculer les équivalences en fonction du niveau de pratique lorsque la distance courue pendant 12 minutes est comprise entre 1500 m et 3000 m ce qui est le cas ici. Si D était nettement supérieur à 3000 m, l'équation B ne pourrait convenir.

Équivalences entre les résultats au test de Cooper et temps sur 2000 m.

	Centile 1 (équation A)		Centile 50 (équation B)		Centile 99 (équation B)	
	Cooper	T2000 mn, s	Cooper	T2000 mn, s	Cooper	T2000 mn, s
Femmes	1512	16.41	2076	11.40	2864	8.12
Hommes	1706	14.49	2414	10.11	3121	7.05*

* Équivalence calculée par extrapolation parce que 3121 est proche de 3000.

Enseignements

* Femmes. Les niveaux médian et très faible indiqués par les tables de Cooper sont donc inférieurs aux niveaux observés chez les candidates femmes en situation de concours CRPE.

Les niveaux supérieurs sont comparables car le traitement des performances amenait à des cotations un peu plus difficiles que celles que nous avons adoptées finalement pour ne pas trop « durcir » le barème 2003.

	C01 et note 1.5	C50 et note 7.5	C99 et note 13.5
Tables de Cooper	16.41	11.40	8.12
Résultats au CRPE	14.37	10.40	8.24

* Hommes. Les niveaux indiqués par les tables de Cooper sont tous nettement inférieurs aux niveaux observés chez les candidats hommes en situation de concours CRPE.

	C01 et note 1.5	C50 et note 7.5	C99 et note 13.5
Tables de Cooper	14.49	10.11	7.05
Résultats au CRPE	10.13	8.00	6.43

Pourquoi ce décalage plus important pour les hommes ? Les candidats hommes au CRPE seraient d'un niveau nettement meilleur que les candidates femmes. Mais on peut aussi émettre l'hypothèse que les femmes les plus faibles dans les populations non entraînées se soumettent plus rarement au test de Cooper que les hommes les plus faibles dans ces populations. Les tables de Cooper prendraient alors plus en compte les performances les plus faibles chez les hommes.

Au-delà de cette hypothèse (à vérifier), on peut dire que les tables qui sont proposées après étude des résultats de 837 candidates et 198 candidats correspondent à des exigences plus élevées que celles qu'on définirait à partir des statistiques de l'Institut pour la recherche en aérobic de Dallas. Ceci peut être dû, entre autres, à des motivations et préparations plus élevées dans une épreuve de concours.

Comparaison avec des résultats de lycéens

Au cours des dernières années, la maîtrise d'exécution ayant un poids très important dans la notation, on a assisté à une régression des performances scolaires : ainsi beaucoup d'élèves cherchaient plus à courir régulièrement qu'à améliorer leurs performances de type sportif. Il est donc très difficile de se fonder sur des statistiques récentes.

Nous nous sommes donc référés aux résultats que nous avons observés sur 1000 m en 1982-84 lors d'évaluations en EPS dans une dizaine d'établissements (mémoire INSEP, 1984). On observait à l'époque que pour les filles les résultats régresaient à partir de 13 ans mais étaient assez voisins entre 16 et 18 ans. Pour les garçons ils progressaient jusqu'à 16 ans et se stabilisaient ensuite. Les niveaux étaient donc proches entre 16 et 19 ans en demi-fond (alors qu'ils variaient fortement avec l'âge au lancer du poids par exemple).

Nous avons donc pu prendre en compte 383 performances pour les filles et 315 pour les garçons réalisés en situation d'évaluation scolaire (abandons non pris en compte), donc dans des conditions de motivation inférieures à celle d'un concours.

Les correspondances des temps sur 1000 m et sur 2000 m ont été calculées selon diverses sources dont la table Letessier. On obtient alors les résultats suivants.

* Test de Cooper, lycéennes et candidates au CRPE.

Filles	note = 0	C01 et n=1.5	C50 et n=7.5	C99 et n=13.5	note =20
D'après tables Cooper		16.41	11.40	8.12	
383 lycéennes de 16 à 19 ans	16.42	15.29	11.42	9.16	8.48
Barème CRPE 2003	13.00	12.40	10.43	9.06	8.50
Propositions CRPE 2004*	16.00	14.37	10.40	8.24	8.00

(* en fonction des performances observées en juin 2003)

Pour le niveau très faible, les résultats sont convergents pour situer les performances bien au-delà des 13 minutes du barème CRPE 2003. Compte tenu de la motivation liée à un concours, un temps de 16 minutes semble acceptable pour le niveau 0.

Pour le niveau médian, les résultats convergent d'avantage avec la cotation du barème 2003 qui peut être conservé.

Pour le niveau supérieur, le test de Cooper et les résultats du CRPE convergent pour situer les exigences au-delà du barème 2003 et de nos estimations à partir de performances de lycéennes. 8 minutes semble accessible puisque 6 candidates sur 837 ont nettement dépassé cette performance.

* Test de Cooper, lycéens, résultats au CRPE.

Garçons	note = 0	C01 et n=1.5	C50 et n=7.5	C99 et n=13.5	note =20
D'après tables Cooper		14.49	10.11	7.05	
315 lycéens de 16-19 ans	12.34	11.35	8.31	6.59	6.43
Barème CRPE 2003	11.00	10.40	8.40	7.05	6.50
Propositions CRPE 2004*	11.00	10.13	8.00	6.43	6.30

(* en fonction des performances observées en juin 2003)

Pour le niveau très faible, les résultats au CRPE sont meilleurs que ceux qu'on pouvait supposer en fonction du Cooper et du niveau lycéen, ce qui paraît normal compte tenu des conditions du concours. Barème CRPE 2003 et barème en fonction des résultats 2003 convergent pour situer la performance aux alentours de 11 minutes.

Pour le niveau médian, les résultats du CRPE sont bien supérieurs aux estimations issus du test de Cooper. Ils sont sensiblement meilleurs que les estimations à partir de performances de lycéens. Le barème 2004 peut être plus exigeant.

Pour le niveau supérieur, le niveau indiqué par les barèmes CRPE 2003 est nettement dépassé par certains candidats alors que les règles 2003 imposaient de rester « dans les limites du barème ». En revanche l'estimation liée au Cooper est bien en-deçà des résultats observés. Une performance située aux alentours de 6.30 paraît une performance accessible.

Annexe 2 : bref mémento statistique

Ce bref mémento ne peut apporter que quelques repères par rapport aux problèmes traités dans cette étude.

Pour plus de précisions on pourra se reporter à Langouët et Porlier, *Mesure et statistique en milieu éducatif* (édition ESF).

Pour construire une table de cotation il faut avoir une représentation aussi précise que possible des performances attendues, en fonction de la réglementation de l'épreuve, du sexe et de l'âge des élèves...

Pour cela on doit souvent réaliser une étude de performances sur un échantillon représentatif de cette population. Cette étude passe par trois phases :

- la construction de l'échantillon et le recueil des données ;
- l'étude descriptive des performances ;
- l'interprétation des résultats en fonction de la population de référence pertinente (locale, nationale... ?) : à partir de l'échantillon peut-on « inférer » les caractéristiques de la « population parente » à laquelle il appartient ?

1. Construire l'échantillon, recueillir les données

Cette phase est capitale. Les conditions dans lesquelles on recueille les performances importent plus que leur nombre. Il s'agit de disposer de « performances représentatives ». Pour cela quelques règles sont à observer :

- Respecter les conditions de déroulement de l'épreuve (matériel, réglementation, motivation...). Recueillir des performances réalisées en situation réelle d'évaluation et non lors d'un entraînement.
- Veiller à la précision des mesures.
- Prendre en compte les performances de différentes classes et différents profs (et de différents établissements si l'on veut dépasser la référence établissement).

Les performances d'une seule classe (14 filles et 12 garçons par exemple) peuvent constituer une base de départ mais ne peuvent être considérées comme un échantillon représentatif car les résultats dépendent trop de conditions spécifiques à la classe (prof, modalités d'enseignement, conditions matérielles, niveau de la classe...). Si l'on veut construire des barèmes propres à un établissement, une soixantaine de performances par sexe et par niveau de classe (ou âge) constitue déjà un ensemble très intéressant.

2. Décrire et résumer les résultats

En possession d'un échantillon de performances, il faut en extraire l'information utile à partir de trois grandes questions :

- Quelle est la forme de la distribution ? Peut-on notamment dire que la distribution est proche d'une distribution gaussienne (ou « normale ») ?
- Quelles sont les tendances centrales (moyenne, médiane, mode...)?
- Comment caractériser la dispersion des données (écart-type, autres indices) ?

La première question est la principale car si la distribution est proche d'une distribution gaussienne (voir plus loin 3.3) on peut résumer la distribution avec seulement deux indices : la moyenne et l'écart type. Si non, il faudra retenir d'autres descripteurs (médiane, centiles, coefficient d'asymétrie...).

2.1. Un exemple sur 14 performances

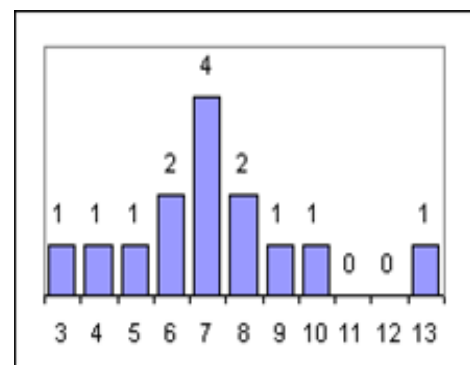
14 filles de 5^e ont réalisé une épreuve d'adresse au basket : marquer le plus de paniers possible en 1' et en tirant d'endroits précis et différents (épreuve de fin de cycle prévue par un collègue en plus de l'évaluation en jeu). On a enregistré les scores suivants.

3	4	5	6	6	7	7	7	7	8	8	9	10	12
---	---	---	---	---	---	---	---	---	---	---	---	----	----

2.2. Distribution des performances et histogramme

Pour décrire un échantillon, une représentation graphique est presque toujours souhaitable. Si les scores observés allaient de 3 à 150 il faudrait créer des classes (de 0 à 10 ; de 11 à 20, etc...). Selon le regroupement en classes la forme de la distribution serait différente.

Scores	Nbre
3	1
4	1
5	1
6	2
7	4
8	2
9	1
10	1
11	0
12	0
13	1



2.3. Quelques descripteurs (valeurs dans l'exemple)

Effectif	On le désigne le plus souvent par N	14
Minimum	valeur observée la plus petite	3
Maximum	Valeur observée la plus grande	13
Étendue	différence entre maximum et minimum.	10
Moyenne arithmétique	somme des données divisée par leur nombre	7.14
Médiane ou centile C50	valeur de la variable telle qu'il y ait autant de données qui lui soient supérieures ou égales que de valeurs qui lui soient inférieures ou égales.	7.0
Mode	valeur de la variable la plus fréquente.	7.0
Écart type	Racine carrée de la variance. Variance = moyenne des carrés des écarts de chaque donnée à la moyenne. ((3 - 7.14) ² + (4 - 7.14) ² ...) / 14. Pour estimer l'écart type de la population parente on divise par N-1 et non par N.	2.42 et 2.51

(Suite page suivante)

Centile C25 (C50–C25)	Valeur de la variable telle que 25% des données lui soient inférieures (calcul par Excel)	6.0 (1.0)
Centile C75 (C75 – C50)	Estimation de la performance telle que 75% des données lui soient inférieures (calcul par Excel)	8.0 (1.0)
Coefficient d'asymétrie utilisé par excel. (il en existe d'autres)	Indice fondé sur les écarts à la moyenne de toutes les données. Si les données supérieures à la moyenne sont plus dispersées que les valeurs inférieures le coefficient est positif. Et inversement.	0.67

Remarque : un centile est une valeur de la variable tel qu'un certain pourcentage des données lui soit inférieur. On retient souvent C10, C20, C30, C40, C50, C60, C70, C80, C90. Ces 9 nombres permettent de résumer des milliers d'observation et de renseigner à la fois sur la forme de la distribution, la dispersion et les tendances centrales (médiane). Les centiles sont très utilisés en statistique descriptive.

2.4. Commentaires sur cet échantillon.

- Un mode se dégage nettement presque au centre de la distribution.
- L'histogramme est symétrique à une donnée près.
- moyenne, médiane et mode sont très proches.
- Le coefficient d'asymétrie est assez élevé (0.67) mais à cause d'un seul score (13). Un autre coefficient d'asymétrie fondé sur l'écart entre la moyenne et la médiane indique une asymétrie très faible. Les centiles C25, C50 et C75 indiquent une symétrie $C50-C25 = C75-C50 = 1.0$ (symétrie).

Sauf pour une donnée extrême, la distribution paraît donc proche d'une distribution gaussienne (voir plus loin). En première approximation on pourrait donc ne retenir que la moyenne et l'écart type pour résumer la distribution.

3. Estimer la population parente

3.1. La représentativité de l'échantillon

Peut-on considérer que cet échantillon est représentatif de la population des performances des filles de 5è dans cette épreuve ?

Par rapport à une référence nationale, la question a peu de sens car les performances dépendent trop des conditions d'enseignement variables d'un collège à l'autre.

Par rapport à la référence locale du collège, la question mérite d'être discutée. Le cycle prévu dans le collège a été réalisé et les performances ont été relevées lors d'une évaluation de fin de cycle. De ces points de vue, les performances sont représentatives. Mais elles émanent d'une seule classe avec un seul enseignant. La diversité de l'établissement n'a pu être respectée et un doute important existe sur la représentativité des performances.

3.2. L'influence des variations aléatoires

Par ailleurs les performances sont peu nombreuses (14), or plus l'effectif est petit et plus les variations aléatoires se font sentir. Dans les mêmes conditions de pratique, avec un autre groupe de 14 filles, il est fort probable que les résultats seraient sensiblement différents. (Ces variations seraient beaucoup moins nettes avec des échantillons de 60 filles).

3.3. La distribution des performances est-elle « normale » ?

Pour discuter de cette question, il faut quelques repères sur la distribution normale. Celle-ci est une distribution théorique étudiée au début du 19è siècle (Gauss, Laplace). La célèbre courbe en cloche de Gauss est une représentation de cette fonction dont l'équation repose sur la moyenne et l'écart type. Cette distribution théorique a plusieurs propriétés :

- moyenne, médiane et mode sont confondus ;
- les distributions normales respectent des fréquences théoriques constantes : ainsi dans toute distribution gaussienne, 68.3% des cas sont compris entre moyenne moins un écart type et moyenne plus un écart type ; 99.7% des cas sont compris entre moyenne – 3 écarts types et moyenne + 3 écarts types. Etc.

Or on constate que certaines distributions empiriques (performances au poids par exemple) sont proches de la loi normale. Dans ce cas, il est très commode d'utiliser le modèle théorique pour rendre compte d'une réalité empirique. Encore faut-il vérifier que l'écart au modèle théorique est suffisamment faible pour être négligé.

Et là, les choses se compliquent car de nombreuses distributions de performances s'écartent nettement du modèle gaussien (demi-fond, natation, épreuves où l'apprentissage technique et la motivation influent fortement sur la performance). Utiliser le modèle gaussien pour résumer les performances aboutit alors à des impasses. (En revanche ce modèle est très utile pour définir une distribution conventionnelle de notes au moment de l'étalonnage d'un barème).

De plus lorsque les distributions de performances paraissent proches de la loi normale, elles diffèrent souvent aux extrémités avec le modèle théorique. Or pour construire un barème on ne peut négliger les extrémités puisque cela concerne les élèves les plus faibles et les plus forts.

Dans notre exemple, tous les tests statistiques déboucheraient sur la conclusion « distribution approximativement normale », ce qui justifierait de se fonder sur la moyenne et l'écart type. Une simulation montre que cette conclusion n'est pas entièrement valable.

3.4. La simulation souvent plus utile que les tests d'inférence statistique

Nous avons vu que lorsque la distribution est gaussienne, 99.7% des cas sont compris entre moyenne – 3 écarts types et moyenne + 3 écarts types. Ici on a

- moyenne – 3 écarts types = -0.4 (score négatif)
- moyenne + 3 écarts types = 14.7

Or un score négatif est impossible dans cette épreuve. La simulation permet donc de voir que moyenne et écart type ne permettent pas de décrire correctement la population parente. Si l'on retient l'hypothèse de normalité, on pourrait rectifier provisoirement : moyenne = 7 et écart type = 2.33. La table de cotation irait alors du score 0 au score 14.

Choix qu'il faudrait ensuite confirmer ou infirmer par une étude plus conséquente sur d'autres classes. Mais l'enseignant peut ainsi construire un premier barème.

Ici on aurait $n = 12 + 3(p - 7) / 2.33$ (n = note, p = performance), s'il a convenu d'étalonner des barèmes avec moyenne des notes = 12 et écart type des notes = 3.