

# QCM à points négatifs au début de l'enseignement supérieur : échecs abusifs et biais de genre.

Jean-Marc Braibant, UCLouvain

François-Marie Gerard, Biefor

Etienne Billat, UCLouvain

jean-marc.braibant@uclouvain.be

## Résumé

Les QCM à points négatifs sont utilisés pour réduire l'effet du hasard dans les réponses données. Pour des raisons fortuites, le même examen a été présenté à deux groupes d'étudiant·es en annonçant deux modalités de correction différentes : l'une sans pénalités, l'autre avec pénalités. Alors que l'équivalence entre les deux groupes a été attestée, les échecs sont plus nombreux dans le groupe « avec pénalités ». L'analyse de ces échecs abusifs montre qu'ils concernent davantage les étudiantes que les étudiants. Nous obtenons ici des résultats tangibles qui démontrent l'existence d'un biais de genre. Celui-ci s'explique par des stratégies de réponses différentes entre les hommes et les femmes. Sur la base de cette étude, le recours à des QCM à points négatifs dans les évaluations certificatives ou dans les examens d'admission de nos étudiant·es ne devrait plus être autorisé.

## Abstract

Negative point MCQs are used to minimize the effect of chance in the answers given. Due to fortuitous reasons, the same exam was presented to two groups of students with two different marking schemes announced: one without penalties, the other with penalties. While equivalence between the two groups was attested, there were more failures in the "with penalties" group. The analysis of these abusive failures shows that they concern more female students than male students. Here we have tangible results that demonstrate the existence of a gender bias. This can be explained by the different response strategies of men and women. On basis of this study, the use of MCQs with negative points in the certification assessments or entrance examinations of our students should no longer be permitted.

## Mots-clés

Evaluation, Réussite éducative, Equité, QCM, Biais de genre.

## 1. Introduction

Le recours à des points négatifs est utilisé dans les QCM. Le principe est d'attribuer des points aux réponses correctes, de ne pas pénaliser les omissions et de retirer des points en cas de réponse incorrecte. Ce barème a régulièrement été remis en question. Une des principales critiques porte sur le fait qu'il laisse une trop grande place aux considérations stratégiques sans lien avec la maîtrise de la matière. De nombreux travaux confirment que la pénalisation des mauvaises réponses n'offre pas d'avantages en termes de fidélité ou de validité de l'examen (Choppin, 1988). D'autres études mettent en évidence un biais de genre à la défaveur des femmes (Baldiga, 2014), mais cette conclusion n'est pas partagée par tous les auteurs. Ce manque de consensus s'explique par le fait que ces travaux ont négligé les différences de paramétrage des tests et/ou ont utilisé des critères d'analyse différents pour des publics présentant des profils académiques diversifiés.

À notre connaissance, il n'existe aucune étude permettant de comparer les performances d'un même échantillon d'étudiants répartis aléatoirement en deux sous-groupes (examen avec ou sans notation négative) dans une situation réelle de certification débouchant sur une note officielle. Pour des raisons d'éthique et d'équité, ce type de design expérimental est inconcevable. Les circonstances nous ont donné l'occasion de mener ce type d'analyse suite à une erreur de communication qui s'est produite à l'UCLouvain lors d'un examen en janvier 2020. Un premier groupe a été informé qu'il s'agissait d'un QCM qui serait noté ainsi : 1 point pour une bonne réponse et 0 point dans tous les autres cas. Les encadreurs du deuxième groupe se sont involontairement trompés : ils ont dit qu'il s'agissait d'un QCM « à points négatifs », c'est-à-dire qu'il y aurait 1 point pour une bonne réponse, 0 point en cas d'abstention (pas de réponse) et -0,25 point pour une mauvaise réponse.

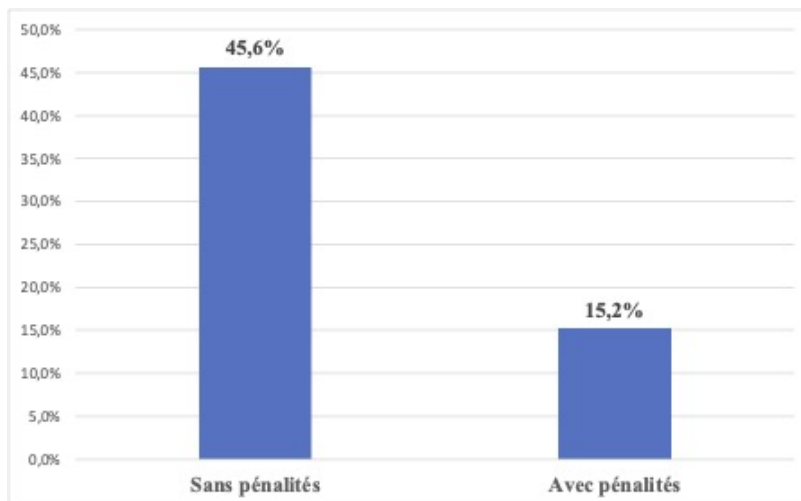
## 2. Objectifs de la recherche

On pourrait qualifier notre recherche d’empirique voire même d’occasionnelle en ce sens que nous avons saisi une « occasion » qui s’est présentée à nous pour répondre à deux questions importantes : quelle est la validité prédictive des deux formes d’examen et qu’en est-il de leur validité de conséquences pour les hommes et pour les femmes ?

## 3. Résultats obtenus dans les deux examens.

L’examen concernait un cours obligatoire donné en 1<sup>ère</sup> année. Il s’adressait à 399 étudiants répartis aléatoirement dans deux amphithéâtres. Après avoir conclu à l’équivalence des deux échantillons, nous montrons que la décision de faire appel ou non à un barème de pénalités a un impact déterminant sur le succès des étudiants (fig. 1). Il y a en effet 3 fois plus de chances d’être en échec dans un examen avec points négatifs.

Figure 1. Pourcentage de réussite selon les 2 conditions



## 4. Analyse de la validité prédictive des deux examens

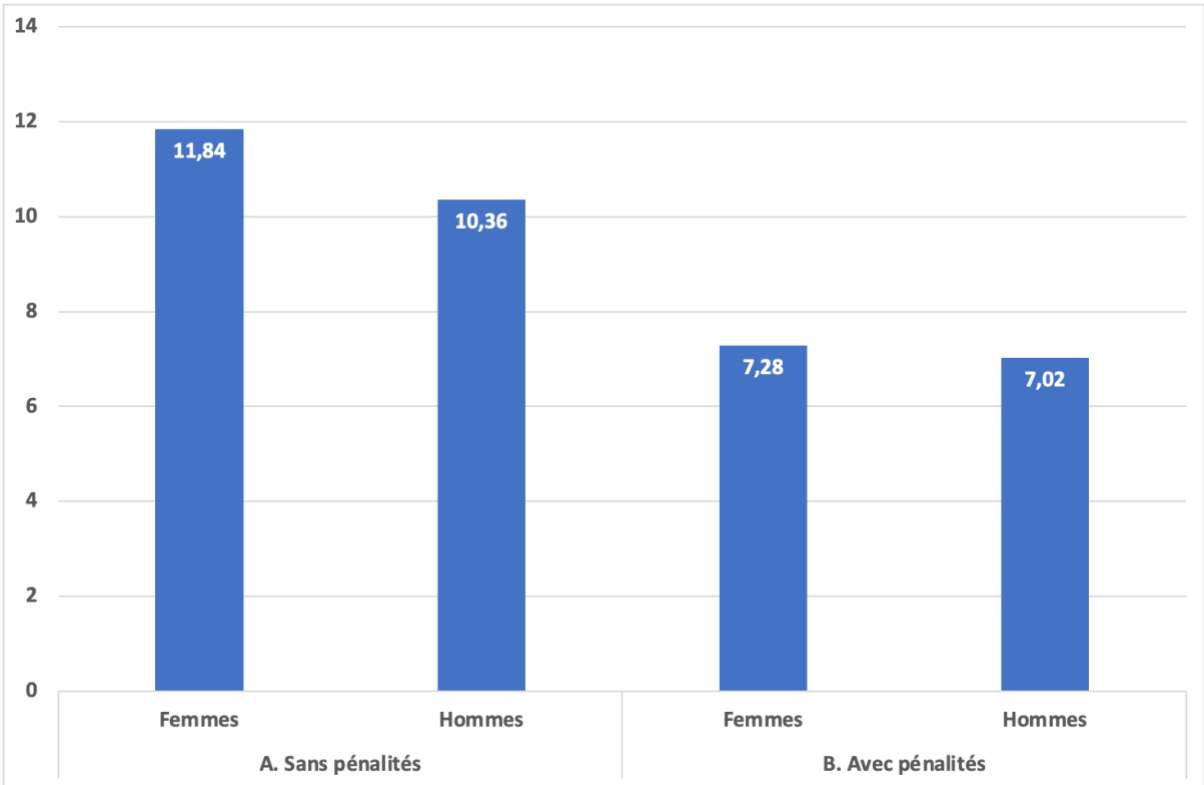
Nous avons mis en relation les notes obtenues par les étudiants à cet examen avec la décision de réussite prise en fin d’année par le jury d’études en utilisant des indicateurs issus du monde médical : la sensibilité d’un test, sa spécificité, le taux de faux positifs et de faux négatifs. Les résultats montrent que l’examen sans pénalités présente de meilleures qualités prédictives que

l'examen avec pénalités. Ce dernier se caractérise par une proportion élevée de « faux négatifs » : on identifie à tort un problème qui n'existe pas pour un nombre conséquent d'étudiants. A l'inverse, l'examen sans points négatifs récompense les étudiants qui ont les meilleures capacités académiques. L'étude des corrélations entre les scores enregistrés dans les deux examens et les résultats obtenus en fin d'année confirme la supériorité prédictive de l'examen sans pénalités.

### 5. Observe-t-on un biais de genre ?

La figure 2 permet de visualiser les performances des étudiantes et des étudiants dans les deux types d'examens. Les différences observées entre les deux conditions sont très significatives tant pour les femmes que pour leurs pairs masculins (t de Student ;  $p < ,001$ )

Figure 2. Moyennes des points dans les 2 situations en fonction du genre

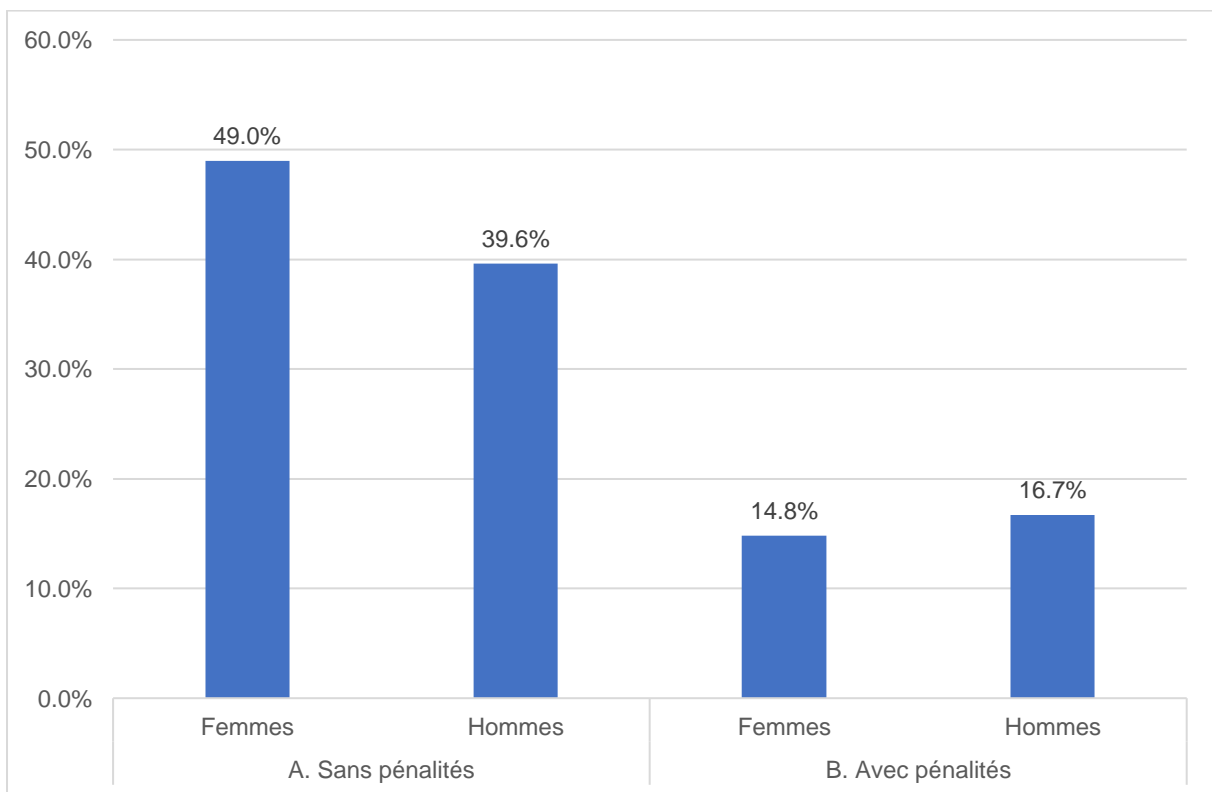


Nous montrons ensuite que les femmes perdent dans l'examen avec pénalités l'avantage qu'elles ont sur les hommes dans l'examen sans pénalités : les différences de performances sont significatives en dessous du seuil de ,05 au test t de Student dans cette dernière condition (p

=,02) alors que ce n'est pas le cas dans l'examen avec pénalités ( $p = ,75$ ). Ce résultat est important, car il met en évidence l'existence d'un biais à la défaveur des femmes.

Les analyses prenant en compte les pourcentages de réussite confirment à quel point le recours à des points négatifs est défavorable aux femmes.

Figure 3. Pourcentage de réussite en fonction du genre dans les deux conditions



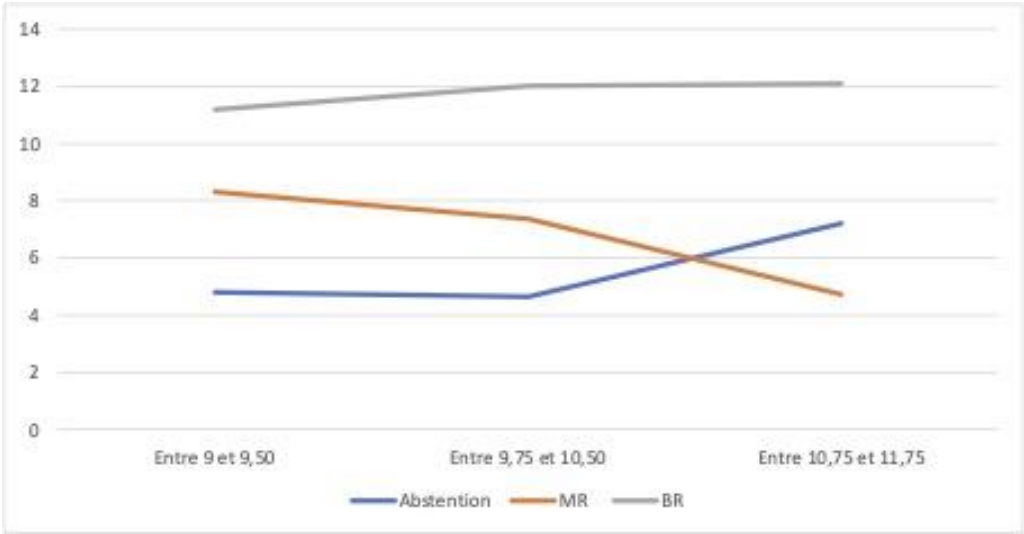
Dans l'examen avec pénalités, les femmes ont un niveau de réussite légèrement inférieur à celui des hommes alors que la condition sans pénalités indique qu'elles ont une meilleure maîtrise de la matière. Leur sous-performance dans l'examen avec pénalités est associée à un risque d'échec très élevé. Ce risque est finalement aussi important que celui des hommes alors que leur connaissance de la matière est meilleure, comme l'indiquent leurs résultats dans l'examen sans pénalités.

## 6. Comment expliquer ce biais de genre ?

Plusieurs études montrent que les femmes ont une plus grande aversion à la prise de risque que les hommes et que cette prudence se traduit dans les QCM à points négatifs par des taux d'abstention plus élevés, ce qui a pour effet de diminuer leur note.

Dans un premier temps, nous avons calculé la moyenne des trois types de réponses possibles en fonction du genre dans les deux types d'examen. Les données montrent que les femmes s'abstiennent un peu plus que les hommes. Cette différence est minime (24 % vs 21 % d'abstentions), mais une étude menée en Belgique montre que tout peut se jouer à un ou à un demi-point près (Roggemans & Spruyt, 2014). Ces auteurs ont notamment suggéré que la propension des femmes à s'abstenir lorsqu'elles ont un niveau proche de la réussite entraînent une bonne partie d'entre elles à l'échec alors que si elles avaient pris le risque de répondre à des questions supplémentaires, elles auraient augmenté sensiblement leurs chances de réussite. C'est ce que nous avons voulu vérifier en nous intéressant aux étudiantes ayant obtenu entre 9 et 11,75 points. Les différents types de réponses qu'elles produisent sont présentés dans la figure 4.

Figure 4. Evolution des trois types de réponses des étudiantes ayant un score proche de la réussite



Les étudiantes les plus proches de la note de 12 sur 24, synonyme de réussite, s'abstiennent davantage que leurs condisciples tout en réduisant sensiblement le nombre de mauvaises réponses. Les deux courbes se croisent. Cette configuration laisse à penser que ces étudiantes adoptent une stratégie d'évitement des risques : elles préfèrent s'abstenir plutôt que de cocher une réponse dont elles ne sont pas certaines. Elles échouent de peu, mais elles échouent, un peu comme ces joueurs de tennis qui retiennent leur coup et n'osent plus monter au filet alors qu'ils

sont sur le point de gagner leur match. Ils finissent par laisser filer la victoire, paralysés par la peur de gagner.

Cette analogie nous a conduit à nous demander s'il n'y avait pas interaction entre la prise de risques et le niveau académique des étudiantes. En d'autres termes, celles qui ont les capacités pour réussir des études universitaires n'auraient-elles pas plutôt tendance à s'abstenir de répondre par rapport à leurs condisciples moins performantes ?

Nous avons testé cette hypothèse en subdivisant les étudiantes ayant échoué à l'examen en fonction de leurs résultats de fin d'année. Les résultats montrent que les étudiantes ayant obtenu moins de 45 crédits s'abstiennent moins de répondre en moyenne que les deux autres catégories. Une analyse des comparaisons des moyennes à l'aide du test t de Student montre qu'il y a une différence significative entre le 1<sup>er</sup> groupe et le 3<sup>e</sup> groupe ( $p = ,02$ ) ainsi qu'entre le 2<sup>e</sup> et le 3<sup>e</sup> groupes ( $p = ,04$ ). En revanche, il n'y a pas de différence entre le 1<sup>er</sup> et le 2<sup>e</sup> groupes ( $p = ,66$ ).

Tableau 1. Evolution des trois types de réponses des étudiantes ayant un score proche de la réussite

	N	Nombre moyen d'abstentions	Écart-type
60 crédits sur 60	40	7,20	5,06
Entre 45 et 60 crédits	51	6,73	5,11
< 45 crédits	34	4,59	4,63

Cette crainte de perdre des points quand on n'est pas totalement certain de la bonne réponse est sans doute à l'origine du sort peu enviable réservé aux étudiantes ayant obtenu 60 crédits sur 60 à la fin de l'année. Elles ne sont en effet que 15 sur 55 à avoir réussi l'examen avec pénalités (27,3 %) alors qu'on enregistre un taux de réussite trois fois supérieur de 77,6 % chez les étudiantes ayant réussi également leur année mais qui ont passé le même examen sans pénalités (45 réussites sur 58).

Ce résultat montre que le biais de genre touche aussi les étudiantes présentant les meilleures habiletés académiques.

## 7. Conclusion

Nos analyses mettent en évidence deux éléments importants. Premièrement, l'examen avec pénalités crée un nombre très élevé d'échecs tout en présentant une validité prédictive moins bonne que l'examen sans pénalités. Deuxièmement, ces échecs abusifs touchent principalement les étudiantes. Compte tenu de leurs résultats dans l'examen sans pénalités, les femmes devraient obtenir de meilleures performances que leurs pairs masculins dans l'examen avec pénalités. Ce n'est pas le cas puisque leur taux de réussite est plus faible que celui des garçons sans pour autant que cette différence soit significative.

Ce type de constat est généralement avancé dans la littérature pour remettre en question l'hypothèse du biais de genre. Puisqu'il n'y a pas de différence significative, il n'y aurait pas non plus de biais. L'intérêt de notre recherche est de montrer que lorsque le niveau de maîtrise de la matière est vérifié par ailleurs, les étudiantes sous-performent de manière importante dans l'examen avec pénalités au point qu'on peut considérer que ce dernier porte atteinte à l'égalité des chances entre les femmes et les hommes.

Sur base de ce travail, le recours à des points négatifs dans les évaluations certificatives ou dans les concours d'admission devrait être proscrit.

### Références bibliographiques

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, 60 (2), 434-448.

Funk, P., & Perrone, H. (2016). *Gender differences in academic performance: The role of negative marking in multiple-choice exams*. Center for Economic and Policy Research Working Paper No DP11716.

Roggemans, L., & Spruyt, B. (2014). *Toelatingsproef (tand)arts : een sociografische schets van de deelnemers en geslaagden*. Vrije Universiteit Brussel.