

Les biais de genre dans les évaluations des enseignements par les étudiant·es à l'université

Une revue de la littérature

Arianna Giorgi et Sara Aguirre Sánchez-Beato



Édition électronique

URL : <https://journals.openedition.org/gef/1309>

ISSN : 2571-7936

Éditeur

Association de recherche sur le genre en éducation et formation (ARGEF)

Référence électronique

Arianna Giorgi et Sara Aguirre Sánchez-Beato, « Les biais de genre dans les évaluations des enseignements par les étudiant·es à l'université », *Genre Éducation Formation* [En ligne], 7 | 2023, mis en ligne le 01 décembre 2023, consulté le 06 janvier 2024. URL : <http://journals.openedition.org/gef/1309>

Ce document a été généré automatiquement le 6 janvier 2024.



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

Les biais de genre dans les évaluations des enseignements par les étudiant·es à l'université

Une revue de la littérature

Arianna Giorgi et Sara Aguirre Sánchez-Beato

1. Introduction

- 1 Dans le monde universitaire, il est courant de recourir aux évaluations des enseignements par les étudiant·es (EEE) en organisant des enquêtes à l'issue de la période d'enseignement : ces questionnaires permettent aux étudiant·es de donner leur avis quant à l'organisation des cours, aux apprentissages acquis et aux compétences des enseignantes. Les EEE constituent un outil important à l'université non seulement en vue d'améliorer la qualité des enseignements, mais aussi en raison du rôle qu'elles jouent dans des pays tels que la Belgique, dans les décisions relatives à l'emploi, à savoir l'embauche, la promotion, le salaire et l'octroi de prix. La question de savoir si les EEE sont biaisées est donc essentielle puisque celles-ci pourraient refléter des stéréotypes, des préjugés et des doubles standards chez les étudiant·es et, par conséquent, conduire à l'application de critères d'évaluation différents pour les enseignants et les enseignantes plutôt que traduire leurs compétences pédagogiques réelles (Basow & Martin, 2012 ; Biernat, 2003 ; Carli, 2001 ; Carli, Alawa, Lee, Zhao & Kim, 2016 ; Eagly & Karau, 2002 ; Foschi, 2000). Dans des milieux traditionnellement masculins, ces biais pourraient aller à l'encontre des enseignantes jusqu'à entraver leur carrière académique : à cause d'évaluations plus négatives de celles des hommes, elles pourraient se voir reléguées à des postes de niveau inférieur, des évaluations négatives étant, dans certains pays, l'un des obstacles principaux à la promotion à l'université. En effet, selon le dernier rapport *She Figures* (Commission Européenne [CE], 2021), en 2018 les femmes représentaient 48,1% des titulaires d'un doctorat, mais leur présence à l'université diminuait au fur et à mesure que l'on avançait dans l'échelle académique,

phénomène dit du « tuyau percé ». Elles occupaient 46,6% des postes universitaires de grade C (postes d'assistant-e, chargée de cours etc.), 40,3% des postes de grade B (postes de professeur-e associé-e) et seulement 26,2% des postes de grade A (postes de professeur-e titulaire). Ainsi, malgré les efforts déployés au niveau européen en matière de recherche et d'innovation, dans le monde académique, les femmes continuent d'occuper majoritairement les échelons les plus bas de la carrière universitaire.

- 2 Cet article vise à identifier si des biais de genre dans les EEE ont été observés dans la littérature scientifique internationale et quelle en est la nature¹. Premièrement, nous illustrerons la méthode de sélection du corpus et celle d'analyse des données. Deuxièmement, nous présenterons les résultats principaux. Enfin, des pistes de réflexion pour ceux et celles qui se consacrent à la recherche dans ce domaine seront proposées. Nous aborderons d'une part la question de la méthodologie employée pour analyser les EEE, de l'autre celle de la culture disciplinaire genrée qui caractérise certains domaines. Nous discuterons ainsi des effets que ces aspects pourraient avoir sur la carrière des femmes à l'université et de l'importance de les prendre en compte pour rendre le monde académique plus égalitaire (notamment en Belgique).

2. Sélection du corpus et méthode d'analyse des données

- 3 Pour réaliser cette revue de la littérature, nous avons d'abord effectué une recherche préliminaire de différents termes en anglais, français et espagnol sur Google Scholar, afin d'identifier les mots-clés principaux employés dans ce type d'études. Ensuite, nous avons identifié les bases de données à partir de trois critères : (1) les domaines disciplinaires qu'elles incluent (2) la langue et le contexte géographique et (3) l'accès à la base de données. Plus précisément, les bases de données devaient être multidisciplinaires ou spécialisées dans les sciences humaines et sociales, inclure des articles en anglais, français et/ou espagnol (langues parlées par les auteures) et permettre le libre accès ou à partir d'un identifiant de notre université, l'Université libre de Bruxelles. Sur base de ces critères, nous avons sélectionné Google Scholar, Scopus, ProQuest Central, Érudit, Scielo et Redalyc. Les trois premières sont des bases de données internationales et multidisciplinaires. Érudit, Scielo et Redalyc sont, au contraire, régionales et de sciences humaines et sociales. Par ailleurs, Érudit est francophone, alors que Scielo et Redalyc publient des articles en espagnol et en portugais.
- 4 Nous avons ensuite effectué une recherche dans les bases de données sélectionnées. Les mots-clés et les opérateurs booléens utilisés peuvent être consultés dans le tableau 1. Chaque algorithme de recherche reposait sur le mot-clé, « genre », combiné avec le mot-clé « évaluations ». Les mots ont été saisis en anglais, français et espagnol. Par ailleurs, en anglais les mots « professor OR faculty OR higher education » ont été ajoutés pour affiner la recherche vu le nombre élevé de résultats.

Tableau 1. Recherche de mots clés

Langue	Mot-clé « genre »	Mot-clé « évaluations »

EN	Gender differences Gender bias Sex differences Sex bias Gender inequality	Student evaluations Student ratings AND (professor OR faculty OR higher education)
FR	Biais de genre Biais de sexe Différences de sexe Différences de genre	Évaluation des enseignements Évaluation de l'enseignement
ES	Sesgos de género Diferencias de género	Evaluación docente Evaluación de la docencia

- 5 Sur base de ces recherches, deux critères ont été appliqués pour présélectionner les documents : le fait qu'ils traitent des EEE dans l'enseignement supérieur et qu'ils examinent les caractéristiques des enseignant-es qui relèvent de l'EDI (égalité, diversité et inclusion). Cette présélection a abouti à un total de 358 documents qui, après l'élimination des doublons, se sont réduits à 245. Les critères d'inclusion ont ensuite été élargis pour comprendre aussi :
- L'accès au texte intégral.
 - La sélection de chapitres de livres et d'articles évalués par les pairs.
 - La sélection de documents de type théorique, empirique, de revue de la littérature ou de méta-analyse.
 - L'application d'une limite de temps aux études empiriques, qui ne devaient pas dater d'avant 2005.
 - L'inclusion uniquement des revues de la littérature systématiques et des méta-analyses dont le sujet principal est le « biais ».
- 6 Après l'application de ces critères, la sélection finale a abouti à un total de 88 documents. Une première lecture des résumés des articles retenus a, par la suite, été menée afin de classer les articles selon :
- Le type d'article : études empiriques, articles théoriques, revues de la littérature et méta-analyses.
 - Le type de données : articles traitant des scores et des commentaires écrits.
 - La méthodologie utilisée pour recueillir les données : études observationnelles (réalisées à partir de l'observation de la réalité et analysant des variables ne pouvant pas être manipulées par les chercheur-es) et études expérimentales ou quasi-expérimentales (où les chercheur-es peuvent créer des situations artificielles, intervenant sur les variables étudiées).
- 7 En raison de la multitude de publications sur le sujet, nous avons choisi de nous concentrer sur les études empiriques observationnelles portant sur les scores, ce qui réduit le corpus des documents présentés à 50. Parmi ceux-ci, une trentaine sont d'origine états-unienne, tandis que les autres proviennent de divers pays du monde (principalement d'Europe et d'Amérique Latine).

- 8 En ce qui concerne la méthode d'analyse, nous avons employé une analyse de contenu (Andréu Abela, 2000). Nous avons créé un tableau, où chaque ligne correspondait à l'un des articles sélectionnés et chaque colonne à l'une des catégories ci-dessous :
- Type d'étude (distinguant les études observationnelles des études expérimentales) ;
 - Caractéristiques de l'échantillon ;
 - Méthodologie d'analyse des données ;
 - Variables étudiées ;
 - Résultats principaux, à savoir les relations entre les variables et la taille des effets.
- 9 Une fois le tableau rempli avec ces informations, nous y avons ajouté quatre colonnes supplémentaires intitulées « Biais liés au genre de l'enseignant·e », « Biais liés au genre de l'étudiant·e », « Biais liés à la discipline » et « Biais liés au niveau/années d'études » où nous avons noté la présence (« oui »), l'absence (« non ») ou le défaut de mention (« / ») de ces biais pour chaque article. Cela nous a permis d'avoir une vue d'ensemble des résultats principaux.

3. Résultats

- 10 Dans cette section, nous présenterons les résultats que l'analyse de contenu a mis en évidence. La première partie sera consacrée au genre de l'enseignant·e et de l'étudiant·e, ainsi qu'à leurs interactions. La deuxième et troisième partie porteront sur les interactions du genre avec le domaine disciplinaire et le niveau d'études respectivement.

3.1. Le genre de l'enseignant·e, le genre de l'étudiant·e et leurs interactions

- 11 La plupart des études traitant des scores ont montré que les EEE ne sont pas exemptes de biais de genre. Très peu sont d'ailleurs celles qui n'ont détecté aucun biais lié au genre des enseignant·es (Acevedo Álvarez & Mairena Rodríguez, 2006 ; Bassey & Joshua, 2016 ; Brockx, Spooren & Mortelmans, 2011 ; Stuber, Watson, Carle & Staggs, 2009 ; Thawabieh, 2017). Néanmoins, ces biais révèlent une grande variété de résultats qui sont rarement convergents, voire contradictoires. Pour rendre plus claire leur complexité, nous distinguerons d'abord entre les études qui montrent des biais en faveur des enseignants (et donc au détriment des enseignantes) et celles qui, au contraire, montrent des biais en faveur des enseignantes (au détriment des enseignants).
- 12 Les premières sont plus nombreuses : une vingtaine d'articles montrent en effet qu'en général les hommes sont évalués plus positivement que les femmes (Adibifar, 2019 ; Al-Maamari, 2015 ; Arámburo Vizcarra & Luna Serrano, 2013 ; Boehmer & Wood, 2017 ; Boring, 2017 ; Boring, Ottoboni & Stark 2016 ; Campbell, Steiner & Gerdes, 2005 ; Fan et al., 2019 ; Fassiotto, Li, Maldonado & Kothary, 2018 ; Fisher, Stinson & Kalajdzic, 2019 ; Johnson, Narayanan & Sawaya, 2013 ; Johnson & Crews, 2013 ; Martin, 2016 ; McPherson, Jewell & Kim, 2009 ; Mengel, Zölitz & Sauermann, 2019 ; Miles & House, 2015 ; Morgan et al., 2016 ; Pama, Dulla & De Leon, 2013 ; Rivera & Tilcsik, 2019 ; Rosen, 2018 ; Stonebraker & Stone, 2015 ; Wagner, Rieger & Voorvelt, 2016 ; Zare-ee, Mohd Don & Tohidian, 2016). Les articles qui révèlent des biais en faveur des enseignantes, qui

obtiennent des scores plus élevés que les hommes, sont moins nombreux (Basow & Montgomery, 2005 ; García Berro, Colom, Martínez, Sallarés & Roca, 2011 ; Gehrt, Louie & Osland, 2015 ; Kohn & Hatfield, 2006 ; Kozub, 2010 ; Martínez-González, Martínez-Stack, Buquet-Corneto, Sánchez-Mendiola & Díaz-Bravo, 2012 ; Smith, Yoo, Farr, Salmon & Miller, 2007). Néanmoins, dans les deux cas, la taille des effets est faible².

- 13 Certaines études montrent des résultats plus nuancés. Si les enseignants sont mieux évalués dans certains aspects, les enseignantes le sont dans d'autres (Angelo et al., 2019 ; Huebner & Magel, 2015 ; Medel & Asún, 2014 ; Ólafsdóttir, 2018 ; Vázquez Arenas et al., 2018), mais encore une fois, la taille de l'effet est faible. Par exemple, l'étude d'Angelo et al. (2019) illustre ces contrastes. À partir de MedHub, un système qui permet aux stagiaires en médecine d'évaluer leurs enseignant·es, les chercheur·es ont analysé les évaluations d'étudiant·es inscrit·es en chirurgie générale. Les résultats montrent que les enseignantes ont obtenu des scores plus favorables dans les domaines des soins aux patient·es, du professionnalisme (respect des rendez-vous, ponctualité etc.) et de la défense des patient·es au sein du système de santé. Quant aux enseignants, ils ont reçu des scores plus élevés dans la catégorie « savoir médical ». Medel et Asún (2014) montrent également que les étudiant·es de la faculté de sciences sociales de l'Universidad de Chile évaluent plus positivement les hommes en ce qui concerne leur maîtrise de la matière, alors qu'ils et elles attribuent des scores plus élevés aux femmes dans les sphères des habiletés pédagogiques et de la responsabilité. Par ailleurs, d'autres études montrent que si l'on demande aux étudiant·es d'attribuer des caractéristiques à leurs enseignant·es, celles utilisées pour décrire les femmes et les hommes sont très différentes entre elles : les femmes sont accessibles, aimables, compatissantes et attentives aux autres, elles sont souvent considérées comme étant en mesure d'interagir avec leurs étudiant·es et de favoriser l'interaction individuelle enseignant·e-étudiant·e. Au contraire, les hommes sont surtout définis par le *leadership*, ils sont perçus comme ayant une meilleure maîtrise de la matière, comme étant plus compétents, innovants, dynamiques et enthousiastes (Basow, Phelan & Capotosto, 2006 ; Burke, Head-Burgess & Siders, 2017). Ces différences d'appréciation indiquent l'application de critères d'évaluation différents pour les enseignants et pour les enseignantes.
- 14 Les interactions entre le genre de l'enseignant·e et celui de l'étudiant·e jouent aussi un rôle déterminant sur les EEE. Plusieurs articles mettent en évidence des *same-gender biais*, à savoir des biais qui montrent une préférence des étudiants pour leurs enseignants d'un côté, et des étudiantes pour leurs enseignantes de l'autre. Basow et al. (2006) montrent à ce propos que les étudiants sont beaucoup plus enclins à désigner comme meilleur professeur un homme. Au contraire, le pourcentage d'étudiants qui nomment leurs enseignantes comme étant les meilleures qu'ils n'aient jamais eues, correspond seulement à la moitié du pourcentage d'enseignantes qu'ils ont effectivement eues. D'autres études (Boring, 2017 ; Boring et al., 2016 ; Zare-ee et al., 2016) montrent des résultats similaires : tant globalement que dans les différentes dimensions de l'enseignement, les hommes reçoivent des scores plus élevés lorsqu'ils sont évalués par des étudiants. Par exemple, la probabilité qu'un étudiant évalue un professeur comme excellent en termes de satisfaction globale est entre 33% et 41% fois plus élevée que celle que le même étudiant qualifie une enseignante d'excellente (Boring, 2017). Mengel et al. (2019) montrent que les enseignantes reçoivent

systématiquement des scores inférieurs à ceux des hommes et ce biais est dû essentiellement aux évaluations des étudiants.

- 15 La concordance de genre prédit aussi les scores que les étudiantes attribuent à leurs enseignantes. Dans certains cas, elle est moins significative que chez les hommes (Boring, 2017 ; Boring et al., 2016), dans d'autres, c'est plutôt le contraire. Gehrt et al. (2015) expliquent d'ailleurs que les étudiantes semblent préférer les enseignantes de rang inférieur, telles que les assistantes ou les professeurs associées, plutôt que celles de rang supérieur, telles que les professeurs titulaires. Cela pourrait montrer un biais en faveur du corps professoral le plus similaire en termes de genre et d'âge relatif.
- 16 D'autres études (Gupta, Garg & Kumar, 2018 ; Price, Svensson, Borell & Richardson, 2017 ; Young, Rush & Shaw, 2009) ne mettent en lumière aucune tendance systématique des étudiant-es à donner des scores différents aux enseignant-es du même genre. Néanmoins, l'effet de ces interactions est faible.
- 17 L'interaction entre le genre de l'enseignant-e et celui de l'étudiant-e se manifeste également à travers le *cross-gender bias*. Dans ce cas-ci, ce sont les étudiants qui évaluent plus positivement les enseignantes, alors que les étudiantes attribuent des scores plus élevés aux enseignants. Deux études (Gupta et al., 2018 ; Kohn & Hatfield, 2006) mettent en évidence ce phénomène. Plus spécifiquement, Kohn et Hatfield (2006) montrent que ce sont les étudiantes qui évaluent les enseignants de façon significativement plus élevée. Cependant, ces différences entre étudiants et étudiantes se doivent d'être nuancées. En effet, d'un point de vue structurel, peu de différences dans les évaluations globales des étudiant-es en fonction de leur genre peuvent être observées.

3.2. Le domaine disciplinaire

- 18 Le domaine disciplinaire et la faculté jouent aussi un rôle dans les EEE, car les étudiant-es n'évaluent pas toujours de la même manière d'une discipline à l'autre. Plus spécifiquement, ceux et celles qui enseignent dans des domaines STIM³ reçoivent des scores significativement plus bas que les enseignant-es dans d'autres disciplines (Acevedo Álvarez & Mairena Rodríguez, 2006 ; Arámburo Vizcarra & Luna Serrano, 2013 ; Basow & Montgomery, 2005 ; Fisher et al., 2019 ; Rosen, 2018 ; Thawabieh, 2017). Basow et Montgomery (2005) montrent à cet égard que les étudiant-es inscrit-es dans les sciences humaines et sociales attribuent à leurs enseignant-es des scores significativement plus élevés par rapport aux étudiant-es en sciences naturelles. Des résultats similaires ont été mis en évidence par Fisher et al. (2019). Rosen (2018) a également montré que les dix disciplines les mieux classées, rangées selon le score moyen de la qualité globale de l'enseignement, sont des sciences humaines, des arts ou des sciences sociales, alors que parmi les dix disciplines les moins bien classées, la majorité sont des domaines STIM. Malgré ces différences significatives, la taille des effets est faible.
- 19 Des différences apparaissent aussi entre les disciplines traditionnellement masculines et celles féminines. Dans le contexte états-unien, Stuber et al. (2009) identifient les sciences dites dures, les mathématiques, l'ingénierie et le business comme étant des disciplines masculines, les sciences humaines et les arts comme étant féminines et les sciences sociales comme étant neutres sur le plan du genre. Ces catégorisations se réfèrent aux expert-es qui ont enseigné et créé les connaissances dans ces domaines,

ainsi qu'au contenu des programmes et aux styles pédagogiques utilisés. Les résultats montrent que les disciplines dites masculines prédisent des scores de qualité globale de l'enseignement plus faibles, ce qui confirme la tendance des étudiant-es dans les STIM à attribuer à leurs enseignant-es des scores plus négatifs que dans les sciences humaines et sociales (Basow & Montgomery, 2005 ; Stuber et al., 2009). Toutefois, Pama et al. (2013) montrent des résultats qui vont dans le sens contraire : en matière d'efficacité de l'enseignement, les scores moyens de la faculté de technologie industrielle sont significativement plus élevés que ceux de la faculté des arts et des sciences. Bien que cette étude soit la seule exception, les auteur-es n'utilisent pas des méthodes d'analyse des données inédites, ni n'émettent d'hypothèses quant à ces résultats en contradiction avec la tendance générale. Or, Pama et al. (2013) ont mené leur étude dans une école spécialisée en sciences et technologie aux Philippines. D'une part, ces résultats pourraient dépendre de l'orientation des étudiant-es qui y sont inscrit-es qui apprécient davantage les enseignant-es des domaines STIM par rapport à ceux et celles des arts. De l'autre, nous ne pouvons pas exclure que ces différences soient de nature culturelle. La dimension genrée des disciplines scientifiques pourrait varier aussi en fonction de l'histoire et de la culture propres aux domaines disciplinaires dans différentes zones géographiques du monde.

- 20 Le domaine disciplinaire peut aussi interagir avec le genre de l'enseignant-e (Basow & Montgomery, 2005 ; Bianchini, Lissoni & Pezzoni, 2013 ; Fassiotto et al., 2018 ; Fisher et al., 2019 ; Gupta et al., 2018 ; Martínez-González et al., 2012 ; Narayanan, Sawaya & Johnson, 2014 ; Price et al., 2017 ; Rivera & Tilcsik, 2019 ; Rosen, 2018). Dans les départements d'ingénierie, d'informatique, de business et économie et de droit, ce sont les enseignantes qui perdent le plus de points. D'ailleurs, si toutes les enseignant-es sont moins bien évaluées, ce sont les femmes qui sont les plus touchées. Dans les autres domaines, en revanche, à savoir l'anglais, la philosophie, l'histoire et l'histoire de l'art, le genre de l'enseignant-e ne semble pas interagir avec la discipline. Néanmoins, la taille des effets de cette interaction est faible.
- 21 Par ailleurs, il semblerait que dans les domaines disciplinaires à faible représentation féminine au sein du corps enseignant, les enseignantes soient moins bien évaluées que leurs homologues masculins. Par contre, dans les domaines à représentation féminine moyenne ou élevée, il n'y aurait pas de différence majeure entre les genres ni de conséquences négatives pour les hommes. Cela montre un effet négatif de la faible représentation des femmes dans certaines disciplines (Fassiotto et al., 2018 ; Narayanan et al., 2014 ; Rivera & Tilcsik, 2019). Les résultats de Martínez-González et al. (2012) sont similaires, mais dans ce cas-ci la sous-représentation des hommes a aussi un effet négatif sur les évaluations, montrant que la faible représentation peut défavoriser autant les hommes que les femmes. Quant à Mengel et al. (2019), leur étude met en exergue une taille des effets assez comparable qui va dans le même sens autant pour les cours avec une majorité d'institutrices que pour ceux avec une majorité d'instituteurs. Cela suggère que les biais au détriment des enseignantes vont à l'encontre des enseignantes en tant que telles plutôt que contre leur faible niveau de représentation.
- 22 Toutefois, Gupta et al. (2018) révèlent un comportement de genre atypique de la part des étudiant-es, qui ont tendance à donner une note plus élevée aux enseignants dans les disciplines à faible représentation masculine et aux enseignantes dans les cours à faible représentation féminine. Si un comportement de genre typique est observé en ingénierie électrique, où les hommes sont mieux évalués que les femmes, en ingénierie

informatique, en ingénierie civile et en mathématiques c'est plutôt le contraire. Malgré la sous-représentation féminine dans ces domaines, les enseignantes ont été mieux notées que leurs homologues hommes, alors que ces derniers ont reçu de meilleurs scores en sciences humaines et sociales où ils sont sous-représentés par rapport aux enseignantes (Gupta et al., 2018). Price et al. (2017) illustrent aussi un comportement de genre atypique : les étudiant-es en informatique attribuent des scores plus élevés aux enseignantes, alors que les étudiant-es en ingénierie environnementale font de même pour leurs enseignants. Ces résultats sont remarquables car au vu de la proportion d'étudiants et de celle d'étudiantes dans ces facultés, l'informatique est davantage masculine, alors que l'ingénierie environnementale est plutôt féminine. Néanmoins, la taille des effets de ces interactions est faible.

- 23 Enfin, certaines études révèlent aussi une interaction entre la discipline, le genre de l'enseignant-e et celui de l'étudiant-e (Gupta et al., 2018 ; Martínez-González et al., 2012 ; Mengel et al., 2019). Mengel et al. (2019) montrent que les étudiants attribuent aux enseignantes des scores inférieurs d'environ 17% à l'écart type des enseignants dans les cours sans contenu mathématique. Quant aux étudiantes, la différence n'est que de 4% et n'est pas significative. Cependant, pour les cours de mathématique, les différences sont plus importantes, car les étudiants attribuent aux enseignantes des scores inférieurs d'environ 32% à l'écart type de leurs homologues hommes. Pour les étudiantes, l'effet est également important : dans les cours de mathématiques, elles attribuent aux enseignantes des scores inférieurs d'environ 28% à l'écart type des enseignants. Par contre, Gupta et al. (2018) et Martínez-González et al. (2012) ne mettent en lumière aucune tendance systématique. Les effets de cette interaction sont faibles.

3.3. Le niveau d'études

- 24 Les biais rencontrés dans les EEE dépendent aussi du niveau d'études ou de l'année d'inscription à l'université (Acevedo Álvarez & Mairena Rodríguez, 2006 ; Arámburo Vizcarra & Luna Serrano, 2013 ; Bianchini et al., 2013 ; Johnson et al., 2013 ; Kozub, 2010 ; McPherson et al., 2009 ; Medel & Asún, 2014 ; Mengel et al., 2019 ; Narayanan et al., 2014 ; Sena & Crable, 2017 ; Vázquez Arenas et al., 2018).
- 25 Les étudiant-es de niveau supérieur (master ou dernière année de bachelier⁴) attribuent des scores plus élevés à leurs enseignant-es que les étudiant-es de premier cycle⁵ (Bianchini et al., 2013 ; Johnson et al., 2013 ; Kozub, 2010). À ce propos, Johnson et al. (2013) mettent en évidence un effet moyen du niveau d'études sur les EEE, alors que les résultats de Kozub (2010) montrent des effets faibles. Seules les recherches d'Acevedo Álvarez et Mairena Rodríguez (2006) montrent le contraire, révélant que plus le nombre d'années passées à l'université est élevé, moins les étudiant-es sont satisfait-es des enseignant-es. Néanmoins, la taille des effets est faible.
- 26 Concernant les interactions du niveau d'études avec le genre de l'enseignant-e, Johnson et al. (2013) montrent que les enseignantes obtiennent des scores significativement plus faibles dans les cours de niveau inférieur (première et deuxième année de premier cycle) que leurs homologues hommes, alors que ceci n'est pas le cas dans les cours de niveau supérieur ou de deuxième cycle, où il n'y a pas de différence entre hommes et femmes. L'effet de cette interaction est néanmoins faible. En revanche, Vázquez Arenas et al. (2018) montrent que les femmes obtiennent des scores moyens légèrement plus

élevés de la part des étudiant-es de niveau master et dans les cours supérieurs de premier cycle, tandis que les hommes obtiennent de meilleures évaluations dans les autres niveaux. Toutefois, ces différences sont petites et peu significatives. Medel et Asún (2014) indiquent que, contrairement à ce qui se passe avec les enseignants, les enseignantes sont moins bien évaluées par les élèves du deuxième cycle et mieux évaluées par les élèves du premier cycle. Cet effet d'interaction n'est pas sans importance, puisque, dans l'ensemble, il fait perdre aux enseignantes environ quatre points en pourcentage, alors que les enseignants gagnent à peu près autant, ce qui produit un écart entre hommes et femmes d'environ huit points.

- 27 Pour finir, Medel et Asún (2014) montrent l'existence d'une interaction entre le niveau d'études, le genre de l'enseignant-e et celui de l'étudiant-e. Celle-ci suit le même schéma que l'interaction ci-dessus, mais la différence entre le premier et le deuxième cycle est encore plus prononcée lorsque les évaluations sont complétées par des étudiants. Autrement dit, la baisse dans les scores attribués aux enseignantes lors du passage du premier au deuxième cycle est presque le double dans les évaluations des étudiants par rapport à celles des étudiantes. Mengel et al. (2019) relèvent aussi une interaction entre ces trois variables : les biais à l'encontre des enseignantes de la part des étudiants sont plus petits lorsqu'ils entrent à l'université en première année de bachelier et environ deux fois plus importants pour les années consécutives. Pour les étudiantes, seules celles en master attribuent des scores plus faibles à leurs enseignantes.

4. Discussion et conclusions

- 28 Les résultats mettent en évidence la complexité des biais qui affectent les EEE et la difficulté d'avoir accès à un instrument fiable reflétant les compétences pédagogiques et les capacités réelles des enseignantes.
- 29 La littérature révèle des biais qui, le plus souvent, défavorisent les enseignantes. Ce désavantage s'accroît dans les domaines STIM et les cours de premier cycle, mais la taille des effets est, dans la majorité des cas, faible. Ces biais sont imputables autant au genre de l'enseignant-e qu'à d'autres variables, telles que le genre de l'étudiant-e, le domaine disciplinaire et le niveau d'études qui, interagissant avec celui-ci, complexifient la manière dont ils se répercutent sur les scores attribués aux enseignantes.
- 30 En plus du genre, les sources de biais potentielles sont très nombreuses. Les études observationnelles montrent d'autres variables qui ne sont pas analysées dans le cadre de cette revue mais qui ont néanmoins un impact sur les EEE. Celles-ci peuvent être classées en trois groupes : les caractéristiques du cours, telles que la taille de la classe (Acevedo Álvarez & Mairena Rodríguez, 2006 ; Arámburo Vizcarra & Luna Serrano, 2013 ; García Berro et al., 2011 ; Johnson et al., 2013 ; Miles & House, 2015 ; Narayanan et al., 2014 ; Ólafsdóttir, 2018), les caractéristiques de l'enseignant-e et les caractéristiques de l'étudiant-e, à savoir l'âge, la race et l'ethnicité (Adibifar, 2019 ; Campbell et al., 2005 ; Johnson & Crews, 2013 ; McPherson et al., 2009 ; Reid, 2010 ; Smith & Hawkins, 2011 ; Smith & Johnson-Bailey, 2011), ou encore la note attendue ou obtenue par les étudiant-es pour un cours (Acevedo Álvarez & Mairena Rodríguez, 2006 ; Boehmer & Wood, 2017 ; Boring, 2017 ; Boring et al., 2016 ; Brockx et al., 2011 ; Campbell et al., 2005 ; García Berro et al., 2011 ; Johnson et al., 2013 ; McPherson et al., 2009 ; Miles & House, 2015 ; Narayanan et al., 2014 ; Thawabieh, 2017). Par ailleurs, les études

expérimentales indiquent qu'il existe aussi des variables, telles que le style d'enseignement (Anderson & Smith, 2005 ; Anderson, 2010 ; Burnell, Cojuharenco & Murad, 2018), qui influencent les différences repérées en termes de scores entre les enseignants et les enseignantes. De telles variables ne peuvent pas être étudiées dans le cadre d'études observationnelles, car elles nécessiteraient l'intervention des chercheur·es et la création de situations artificielles pour vérifier leur effet sur les résultats, toutes autres conditions étant égales par ailleurs. Toutefois, elles pourraient affecter et complexifier la manière des étudiant·es d'évaluer les enseignant·es.

- 31 Ensuite, les scores attribués aux enseignant·es sont loin d'avoir toujours la même signification. Cette revue permet de prendre conscience de la difficulté de comprendre la réelle signification des scores ainsi que de comparer les EEE entre elles. Cette difficulté découle non seulement du fait que les questionnaires d'évaluation ne sont guère standardisés d'une université à l'autre, mais elle est également observable au sein d'une même institution. La littérature illustre à quel point la façon d'évaluer les enseignant·es peut changer en fonction du genre de l'enseignant·e qui donne le cours, du genre de l'étudiant·e qui l'évalue, du domaine disciplinaire auquel l'étudiant·e appartient et de son niveau d'études. Ainsi, le fait d'avoir recours à un même instrument d'évaluation ne rend pas les EEE forcément comparables, puisque d'un groupe d'étudiant·es à l'autre la signification attribuée aux scores peut changer considérablement.
- 32 La question méthodologique ne doit pas être sous-estimée non plus. L'ampleur de l'impact des biais pourrait découler, entre autres, de l'étendue de l'échelle employée dans les EEE. Rivera et Tilcsik (2019) montrent que si sous l'échelle de dix points les enseignantes sont moins bien évaluées que les enseignants, sous l'échelle de six points et surtout dans les matières à prédominance masculine, les différences entre les évaluations des hommes et des femmes disparaissent ou diminuent considérablement. Autrement dit, l'impact des biais pourrait être lié à un artefact de la méthodologie employée pour recueillir les données ainsi que pour les analyser. La difficulté à repérer des biais de genre dans certaines études semblerait plutôt liée à une question méthodologique : il est d'ailleurs difficile d'imaginer que le monde académique échappe à une dynamique de genre et que les étudiant·es n'aient pas des attentes genrées vis-à-vis des enseignant·es. Dans le monde académique, les hommes continuent d'être la norme, le point de référence duquel les femmes s'écartent : ils sont perçus comme plus compétents et comme beaucoup plus influents que les enseignantes (Carli, 2001 ; Laube, Massoni, Sprague & Ferber, 2007 ; Sprague & Massoni, 2005). Il faut donc analyser les EEE avec prudence puisque certains choix méthodologiques, comme les analyses statistiques des scores, pourraient rendre invisibles des nuances que l'analyse qualitative pourrait par contre relever. D'autre part, la présence de différences ne conduit pas automatiquement à des biais. Les écarts observés entre les scores des enseignants et des enseignantes pourraient être également le résultat de différences réelles dans la manière de donner cours des hommes et des femmes. Pour ces raisons, il serait souhaitable non seulement de faire appel à une analyse statistique des scores, mais aussi d'élargir le cadre d'analyse, mettant en exergue ce que les commentaires écrits et les réponses à des questions ouvertes pourraient apporter. En raison des avantages de l'analyse des commentaires, il est de plus en plus courant d'y avoir recours (Heath et al., 2019 ; Rojek et al., 2019 ; Terkik, Prud'hommeaux, Alm, Homan & Franklin, 2016 ; Wallace, Lewis & Allen, 2019).

- 33 Par ailleurs, définir comme des biais les différences dans la manière des étudiant-es d'évaluer les enseignants et les enseignantes part du présupposé qu'il est possible de mesurer des écarts par rapport à une manière de juger non biaisée. Toutefois, cette vision ne s'intéresse pas à l'activité de jugement en elle-même, ni à la façon dont les acteur-trices sociaux-ales, ici les étudiant-es, produisent leurs jugements. Les différences spécifiques propres aux étudiant-es dans la manière de construire leur jugement constituent un aspect intéressant à approfondir qui, une fois de plus, représente une raison de se méfier des EEE (Bailey, Gupta & Schrader, 2000), et d'une analyse focalisée exclusivement sur les scores.
- 34 Nous aimerions revenir aussi sur les interactions entre le genre de l'enseignant-e et le domaine disciplinaire et surtout sur le concept de culture disciplinaire. La majorité de la littérature scientifique a relevé des biais à l'encontre des enseignantes dans les domaines STIM. Ces résultats ne sont pas étonnants si l'on considère une certaine histoire des sciences qui en fait une activité masculine. À ce titre elle apparaît liée à des valeurs et normes masculines, telles que l'objectivité, l'ordre et la hiérarchie, au détriment des celles jugées (ou construites) plus féminines comme la subjectivité et la spontanéité. Les activités jugées féminines, tout en s'appuyant sur un savoir-faire spécifique, ont été historiquement exclues du domaine des sciences pour être classées comme des arts (Bernardis, 2013 ; Carli et al., 2016 ; Fox Keller, 1985, 1991 ; Ginzberg, 1987 ; Löwy, 1995). Intégrer une perspective de genre dans l'histoire sociale et culturelle des sciences permet déjà d'éclairer les mécanismes qui ont servi à établir les contours des disciplines scientifiques. Selon Gilbert, Crettaz Von Roten et Alvarez (2006), l'exclusion des femmes de certains domaines (comme les branches techniques et scientifiques) ne s'expliquerait pas tant par la socialisation différenciée sur le plan sexué qu'à leur culture disciplinaire. Il faut donc dépasser cette perspective de socialisation, resituer l'exclusion féminine dans le cadre de la sociologie des sciences et se focaliser sur les savoirs des différentes disciplines, leurs modes de production, leurs formes de communication et d'organisation, leurs valeurs et normes. À partir de cette perspective, Gilbert et al. (2006) montrent que ce sont la prédominance d'une rationalité en finalité ainsi que le caractère spécialisé du savoir qui font que certains domaines, tels que les STIM, restent moins accessibles aux femmes. Les cultures disciplinaires de ces domaines reproduisent les caractéristiques supposées et construites d'un point de vue sociohistorique des mondes masculins fondés sur la rigueur et l'uniformité et résistent face à la féminisation et la diversification. Ces cultures seraient donc non seulement genrées mais elles auraient, en même temps, un effet de genre propre (Gilbert, 2009). Néanmoins, selon le pays ou la zone géographique, la dimension genrée d'un même domaine présente une certaine variabilité. Les cultures disciplinaires de certains domaines et les effets de genre qui les accompagnent pourraient ne pas être exactement les mêmes d'une région du monde à l'autre, en fonction aussi de leur histoire et culture. Dès lors, lorsqu'on travaille sur les biais de genre dans les EEE, il faut prendre conscience de ces différences afin de mettre en pratique des stratégies efficaces pour rendre le monde académique plus égalitaire.
- 35 La suppression des obstacles à la participation des femmes à la recherche et à l'innovation est l'un des objectifs de l'Espace Européen de la Recherche. Dans ce contexte, la mise en place de plans d'égalité de genre dans les universités est de plus en plus courante. Cet article, seule revue sur le sujet en français, apporte un regard critique sur les EEE dans le milieu universitaire actuel où favoriser l'égalité de genre

dans les carrières scientifiques devient une priorité. Nous espérons contribuer ainsi à la réflexion et au débat sur la pertinence de ces évaluations dans le cadre des politiques de genre.

BIBLIOGRAPHIE

- Acevedo Álvarez, R. & Mairena Rodríguez, N. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria : Un modelo jerárquico lineal. *Archivos Analíticos de Políticas Educativas*, 14(34), 1-22.
- Adibifar, K. (2019). Student Evaluations of Professors : Does A Professors' Gender, Race, or Age Influence Student Ratings? *International Journal of Active Learning*, 4(2), 88-98.
- Al-Maamari, F. (2015). Response Rate and Teaching Effectiveness in Institutional Student Evaluation of Teaching : A Multiple Linear Regression Study. *Higher Education Studies*, 5(6), 9-20.
- Anderson, K. J. (2010). Students' stereotypes of professors : An exploration of the double violations of ethnicity and gender. *Social Psychology of Education*, 13(4), 459-472.
- Anderson, K. J. & Smith, G. (2005). Students' Preconceptions of Professors : Benefits and Barriers According to Ethnicity and Gender. *Hispanic Journal of Behavioral Sciences*, 27(2), 184-201.
- Andréu Abela, J. (2000). Las técnicas de análisis de contenido: una revisión actualizada. *Fundación Centro Estudios Andaluces, Universidad de Granada*, 10(2), 1-34.
- Angelo, J. L., Moazzez, A., Neville, A., Dauphine, C., Lona, Y. & de Virgilio, C. (2019). Investigating Gender Differences in Faculty Evaluations by Trainees in a Gender-Balanced General Surgery Program. *Journal of Surgical Education*, 76(6), e132-e137.
- Arámburo Vizcarra, V. & Luna Serrano, E. (2013). La influencia de las características del profesor y del curso en los puntajes de evaluación docente. *Revista Mexicana de Investigación Educativa*, 18(58), 949-968.
- Bailey, C. D., Gupta, S. & Schrader, R. W. (2000). Do students' judgment models of instructor effectiveness differ by course level, course content, or individual instructor? *Journal of Accounting Education*, 18(1), 15-34.
- Basow, S. A. & Martin, J. L. (2012). Bias in student evaluations. In M. E. Kite (Éd.), *Effective Evaluation of Teaching. A Guide for Faculty and Administrators* (pp. 40-49). Washington : Society for the Teaching of Psychology.
- Basow, S. A. & Montgomery, S. (2005). Student Ratings and Professor Self-Ratings of College Teaching : Effects of Gender and Divisional Affiliation. *Journal of Personnel Evaluation in Education*, 18, 91-106.
- Basow, S. A., Phelan, J. E. & Capotosto, L. (2006). Gender Patterns in College Students' Choices of Their Best and Worst Professors. *Psychology of Women Quarterly*, 30, 25-35.
- Bassey, B. A. & Joshua, M. T. (2016). Interaction effects of lecturers' gender, discipline and rank on their teaching effectiveness. *Global Journal of Educational Research*, 15, 103-112.

- Bernardis, M.-A. (2013). La culture scientifique, une culture au masculin ? *La Lettre de l'OCIM. Musées, Patrimoine et Culture scientifiques et techniques*, 149, 16-24.
- Bianchini, S., Lissoni, F. & Pezzoni, M. (2013). Instructor characteristics and students' evaluation of teaching effectiveness : Evidence from an Italian engineering school. *European Journal of Engineering Education*, 38(1), 38-57.
- Biernat, M. (2003). Toward a Broader View of Social Stereotyping. *American Psychologist*, 58(12), 1019-1027.
- Boehmer, D. M. & Wood, W. C. (2017). Student vs. Faculty perspectives on quality instruction : Gender bias, "hotness," and "easiness" in evaluating teaching. *Journal of Education for Business*, 92(4), 173-178.
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41.
- Boring, A., Ottoboni, K. & Stark, P. B. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*, 1-11.
- Brockx, B., Spooren, P. & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, 23, 289-306.
- Burke, A. S., Head-Burgess, W. & Siders, M. (2017). He's Smart and She's Nice : Student Perceptions of Male and Female Faculty. *International Journal of Gender and Women's Studies*, 5(1), 1-6.
- Burnell, N., Cojuharenco, I. & Murad, Z. (2018). Now you see it now you don't : The effect of teaching style and seniority on gender bias in teaching evaluations. *Better for Less. Discussion Paper Series*, 9.
- Campbell, H. E., Steiner, S. & Gerdes, K. (2005). Student Evaluations of Teaching : How You Teach and Who You Are. *Journal of Public Affairs Education*, 11(3), 211-231.
- Carli, L. L. (2001). Gender and Social Influence. *Journal of Social Issues*, 57(4), 725-741.
- Carli, L. L., Alawa, L., Lee, Y., Zhao, B. & Kim, E. (2016). Stereotypes About Gender and Science: Women ≠ Scientists. *Psychology of Women Quarterly*, 40(2), 244-260.
- Commission Européenne. (2021). *She Figures 2021*. <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/61564e1f-d55e-11eb-895a-01aa75ed71a1>
- Eagly, A. H. & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R. & Johnston, E. L. (2019). Gender and cultural bias in student evaluations : Why representation matters. *PLoS ONE*, 14(2), 1-16.
- Fassiotto, M., Li, J., Maldonado, Y. & Kothary, N. (2018). Female Surgeons as Counter Stereotype : The Impact of Gender Perceptions on Trainee Evaluations of Physician Faculty. *Journal of Surgical Education*, 75(5), 1140-1148.
- Fisher, A. N., Stinson, D. A. & Kalajdzic, A. (2019). Unpacking Backlash : Individual and Contextual Moderators of Bias against Female Professors. *Basic and Applied Social Psychology*, 41(5), 305-325.
- Foschi, M. (2000). Double Standards for Competence: Theory and Research. *Annual Review of Sociology*, 26(1), 21-42.
- Fox Keller, E. (1985). *Reflections on Gender and Science*. Yale University Press.

- Fox Keller, E. (1991). The Wo/Man Scientist : Issues of Sex and Gender in the Pursuit of Science. In H. Zuckerman, J. R. Cole & J. T. Bruer (éds.), *The Outer Circle: Women in the Scientific Community* (pp. 227-236). W. W. Norton & Co.
- García-Berro, E., Colom, X., Martínez, É., Sallarés, J. & Roca, S. (2011). La encuesta al alumnado en la evaluación de la actividad docente del profesorado. *Aula Abierta*, 39(3), 3-14.
- Gehrt, K., Louie, T. A. & Osland, A. (2015). Student and Professor Similarity : Exploring the Effects of Gender and Relative Age. *Journal of Education for Business*, 90(1), 1-9.
- Gilbert, A. (2009). Disciplinary cultures in mechanical engineering and materials science : Gendered/gendering practices? *Equal Opportunities International*, 28(1), 24-35.
- Gilbert, A.-F., Crettaz Von Roten, F. & Alvarez, E. (2006). Le poids des cultures disciplinaires sur le choix d'une formation supérieure technique ou scientifique : Une perspective genre. *Swiss Journal of Sociology*, 32(1), 141-161.
- Ginzberg, R. (1987). Uncovering Gynocentric Science. *Hypatia*, 2(3), 89-105.
- Gupta, A., Garg, D. & Kumar, P. (2018). Analysis of Students' Ratings of Teaching Quality to Understand the Role of Gender and Socio-Economic Diversity in Higher Education. *IEEE Transactions on Education*, 61(4), 319-327.
- Hameed, F., Ali, A., Hameed, A., Saleem, Z. & Javed, Y. (2015). Teacher evaluation : The role of gender. *Quality & Quantity: International Journal of Methodology*, 49(5), 1779-1789.
- Heath, J. K., Weissman, G. E., Clancy, C. B., Shou, H., Farrar, J. T. & Dine, C. J. (2019). Assessment of Gender-Based Linguistic Differences in Physician Trainee Evaluations of Medical Faculty Using Automated Text Mining. *JAMA Network Open*, 2(5), 1-11.
- Huebner, L. & Magel, R. C. (2015). A Gendered Study of Student Ratings of Instruction. *Open Journal of Statistics*, 5(6), 552-567.
- Johnson, M. D., Narayanan, A. & Sawaya, W. J. (2013). Effects of Course and Instructor Characteristics on Student Evaluation of Teaching across a College of Engineering. *Journal of Engineering Education*, 102(2), 289-318.
- Johnson, R. R. & Crews, A. D. (2013). My Professor is Hot ! Correlates of RateMyProfessors.com Ratings for Criminal Justice and Criminology Faculty Members. *American Journal of Criminal Justice*, 38, 639-656.
- Kohn, J. & Hatfield, L. (2006). The role of gender in teaching effectiveness ratings of faculty. *Academy of Educational Leadership Journal*, 10(3), 121-138.
- Kozub, R. M. (2010). Relationship Of Course, Instructor, And Student Characteristics To Dimensions Of Student Ratings Of Teaching Effectiveness In Business Schools. *American Journal of Business Education*, 3(1), 33-40.
- Laube, H., Massoni, K., Sprague, J. & Ferber, A. L. (2007). The Impact of Gender on the Evaluation of Teaching : What We Know and What We Can Do. *NWSA Journal*, 19(3), 87-104.
- Löwy, I. (1995). Le genre dans l'histoire sociale et culturelle des sciences. *Annales. Histoire, Sciences Sociales*, 50(3), 523-529.
- Martin, L. L. (2016). Gender, Teaching Evaluations, and Professional Success in Political Science. *PS: Political Science & Politics*, 49(02), 313-319.

- Martínez-González, A., Martínez-Stack, J., Buquet-Corneto, A., Sánchez-Mendiola, M. & Díaz-Bravo, P. (2012). Satisfacción de los estudiantes de medicina con el desempeño de sus docentes : Género y situaciones de enseñanza. *Investigación en Educación Médica*, 1(2), 64-74.
- McPherson, M. A., Jewell, R. T. & Kim, M. (2009). What Determines Student Evaluation Scores ? A Random Effects Analysis of Undergraduate Economics Classes. *Eastern Economic Journal*, 35(1), 37-51.
- Medel, R. & Asún, R. (2014). Encuestas de evaluación docente y sesgos de género : Un estudio exploratorio. *Calidad en la Educación*, 40, 172-199.
- Mengel, F., Zöhlitz, U. & Sauer mann, J. (2019). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, 17(2), 535-566.
- Miles, P. & House, D. (2015). The Tail Wagging the Dog ; An Overdue Examination of Student Teaching Evaluations. *International Journal of Higher Education*, 4(2), 116-126.
- Morgan, H. K., Purkiss, J. A., Porter, A. C., Lypson, M. L., Santen, S. A., Christner, J. G., Grum, C. M. & Hammoud, M. M. (2016). Student Evaluation of Faculty Physicians : Gender Differences in Teaching Evaluations. *Journal of Women's Health*, 25(5), 453-456.
- Narayanan, A., Sawaya, W. J. & Johnson, M. D. (2014). Analysis of Differences in Nonteaching Factors Influencing Student Evaluation of Teaching between Engineering and Business Classrooms : Nonteaching Factors Influencing Student Evaluation of Teaching. *Decision Sciences Journal of Innovative Education*, 12(3), 233-265.
- Ólafsdóttir, K. (2018). Gender bias in student evaluation of teaching among undergraduate business students. *Tímarit Um Viðskipti Og Efnahagsmál*, 15(1), 75-86.
- Pama, A. B., Dulla, L. B. & De Leon, R. C. (2013). Student Evaluation of Teaching Effectiveness : Does Faculty Profile Really Matter? *Human Behavior, Development and Society*, 8(1), 94-102.
- Price, L., Svensson, I., Borell, J. & Richardson, J. T. E. (2017). The Role of Gender in Students' Ratings of Teaching Quality in Computer Science and Environmental Engineering. *IEEE Transactions on Education*, 60(4), 281-287.
- Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.com. *Journal of Diversity in Higher Education*, 3(3), 137-152.
- Rivera, L. A. & Tilcsik, A. (2019). Scaling Down Inequality : Rating Scales, Gender Bias, and the Architecture of Evaluation. *American Sociological Review*, 84(2), 248-274.
- Rojek, A. E., Khanna, R., Yim, J. W. L., Gardner, R., Lisker, S., Hauer, K. E., Lucey, C. & Sarkar, U. (2019). Differences in Narrative Language in Evaluations of Medical Students by Gender and Under-represented Minority Status. *Journal of General Internal Medicine*, 34(5), 684-691.
- Rosen, A. S. (2018). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching : A large-scale study of RateMyProfessors.com data. *Assessment & Evaluation in Higher Education*, 43(1), 31-44.
- Sena, M. & Crable, E. (2017). RateMyInformationSystemsProfessor : Exploring the factors that influence student ratings. *Information Systems Education Journal*, 15(6), 56-61.
- Smith, B. P. & Hawkins, B. (2011). Examining Student Evaluations of Black College Faculty : Does Race Matter? *The Journal of Negro Education*, 80(2), 149-162.
- Smith, B. P. & Johnson-Bailey, J. (2011). Student Ratings of Teaching Effectiveness : Implications for Non-White Women in the Academy. *The Negro Educational Review*, 62-63(1-4), 115-140.

- Smith, S. W., Yoo, J. H., Farr, A. C., Salmon, C. T. & Miller, V. D. (2007). The Influence of Student Sex and Instructor Sex on Student Ratings of Instructors : Results from a College of Communication. *Women's Studies in Communication*, 30(1), 64-77.
- Sprague, J. & Massoni, K. (2005). Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. *Sex Roles*, 53(11-12), 779-793.
- Stonebraker, R. J. & Stone, G. S. (2015). Too Old to Teach ? The Effect of Age on College and University Professors. *Research in Higher Education*, 56(8), 793-812.
- Stuber, J. M., Watson, A., Carle, A. & Staggs, K. (2009). Gender expectations and on-line evaluations of teaching : Evidence from RateMyProfessors.com. *Teaching in Higher Education*, 14(4), 387-399.
- Terkik, A., Prud'hommeaux, E., Alm, C. O., Homan, C. M. & Franklin, S. (2016). Analyzing Gender Bias in Student Evaluations. *Proceedings of COLING 2016, 26th International Conference of Computational Linguistics: Technical Papers*, 868-876.
- Thawabieh, A. M. (2017). Students Evaluation of Faculty. *International Education Studies*, 10(2), 35-43.
- Vázquez Arenas, G., Conesa Pastor, J., García Cascales, M. S., García-León, J., García Martín, A. & Solano Fernández, J. P. (2018). ¿Existe sesgo de género en la valoración de la actividad docente de títulos de Edificación? *Advances in Building Education / Innovación Educativa en Edificación*, 2(3), 80-93.
- Wagner, N., Rieger, M. & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations : Evidence from mixed teaching teams. *Economics of Education Review*, 54, 79-94.
- Wallace, S. L., Lewis, A. K. & Allen, M. D. (2019). The State of the Literature on Student Evaluations of Teaching and an Exploratory Analysis of Written Comments : Who Benefits Most? *College Teaching*, 67(1), 1-14.
- Young, S., Rush, L. & Shaw, D. (2009). Evaluating Gender Bias in Ratings of University Instructors' Teaching Effectiveness. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-14.
- Zare-ee, A., Mohd Don, Z. & Tohidian, I. (2016). Gender differences in students' ratings of university teachers in the Iranian education system. *Learning and Teaching in Higher Education: Gulf Perspectives*, 13(1), 19-35.

NOTES

1. L'article a été réalisé dans le cadre du Plan Diversité de l'Université libre de Bruxelles dont l'une des mesures s'intéresse aux biais de genre dans les EEE en raison de leurs implications pour la carrière du personnel académique et scientifique. À l'ULB, les notes obtenues dans les EEE sont incluses dans les dossiers de recrutement et de promotion des enseignantes. Par mesure de précaution, les enseignantes peuvent inclure un rapport qui explique leurs évaluations.
2. Nous interprétons la taille des effets à partir des classifications de Cohen (1988, 1992). Les auteur-es emploient les valeurs Eta-carré (η^2) ou Eta-carré partiels (ηp^2) (0.01 = effet faible, 0.06 = effet moyen, 0.14 = effet fort). Parfois les valeurs d de Cohen sont aussi employées (0.2-0.3 = effet faible, 0.5 = effet moyen, à partir de 0.8 = effet fort). Dans quelques cas, le coefficient r de corrélation est utilisé (0.1-0.3 = effet faible, 0.3-0.5 = effet moyen, 0.5 = effet fort).
3. Sciences, technologie, ingénierie et mathématiques.

4. En Belgique, le bachelier (« *bachelor* ») correspond aux premières trois années d'études universitaires.

5. Le premier cycle correspond aux premières années d'études universitaires après les études secondaires. Il est à distinguer du deuxième cycle, ou niveau master, qui permet la spécialisation dans un domaine spécifique.

RÉSUMÉS

Dans certains pays, les évaluations des enseignements par les étudiant-es (EEE) sont considérées dans les décisions relatives à l'emploi à l'université, où la participation des femmes diminue au fur et à mesure que l'on gravit les échelons de la carrière scientifique. S'insérant dans le cadre des politiques d'égalité à l'université, cette revue systématique vise à identifier la présence et la nature des biais de genre dans la littérature sur les EEE. En général les biais repérés défavorisent les enseignantes, surtout en sciences et techniques et dans le premier cycle d'études, mais la taille des effets est faible. Menant une discussion critique, nous réfléchissons à l'importance de comprendre la signification des scores attribués aux enseignant-es par les étudiant-es qui pourrait découler de la méthodologie utilisée pour analyser les EEE et des cultures propres aux domaines disciplinaires. Ces aspects sont essentiels pour élaborer des politiques de genre efficaces.

INDEX

Mots-clés : évaluations des enseignements par les étudiant-es (EEE), biais de genre, enseignement supérieur

AUTEURS

ARIANNA GIORGI

Université libre de Bruxelles

SARA AGUIRRE SÁNCHEZ-BEATO

Université libre de Bruxelles