



# Sekseverschillen in vakevaluaties

Studentbeoordelingen van mannelijke en vrouwelijke  
docenten in UvA Q vakevaluaties

Jeroen van den Tillaart  
DAIR seminar 2018



# Waarom dit onderzoek?

On the credibility of the judge

A cross-classified multilevel analysis on student teaching

Pieter Spooren\*

Department of Social and Political Sciences, University of Antwerp

Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams

Natascha Wagner\*, Matthias Rieger, Ka  
International Institute of Social Studies, Erasmus University Rotterdam, Netherlands

What's in a Name: Exposing Gender Bias in Student Ratings of Teaching

Lillian MacNeill • Adam Driscoll • Andrea N. F.

GENDER BIAS IN TEACHING EVALUATIONS

by

Sauermann & Frey

GUARDING AGAINST  
POTENTIAL BIAS IN  
STUDENT EVALUATIONS

WHAT EVERY FACULTY MEMBER NEEDS TO KNOW

Tamara Baldwin and Nancy Blattner

# Onderzoeksvragen

- Krijgen mannelijke en vrouwelijke docenten gemiddeld dezelfde rapportcijfers in de vakevaluaties? Zo nee, welke kant gaat dit verschil op?
- Zijn er docent-, cursus- en/of surveyeigenschappen aan te wijzen die hiermee samenhangen?
- Hoe staat het ten aanzien van deelaspecten van het doceren waarnaar gevraagd wordt in de vakevaluaties?



# Data

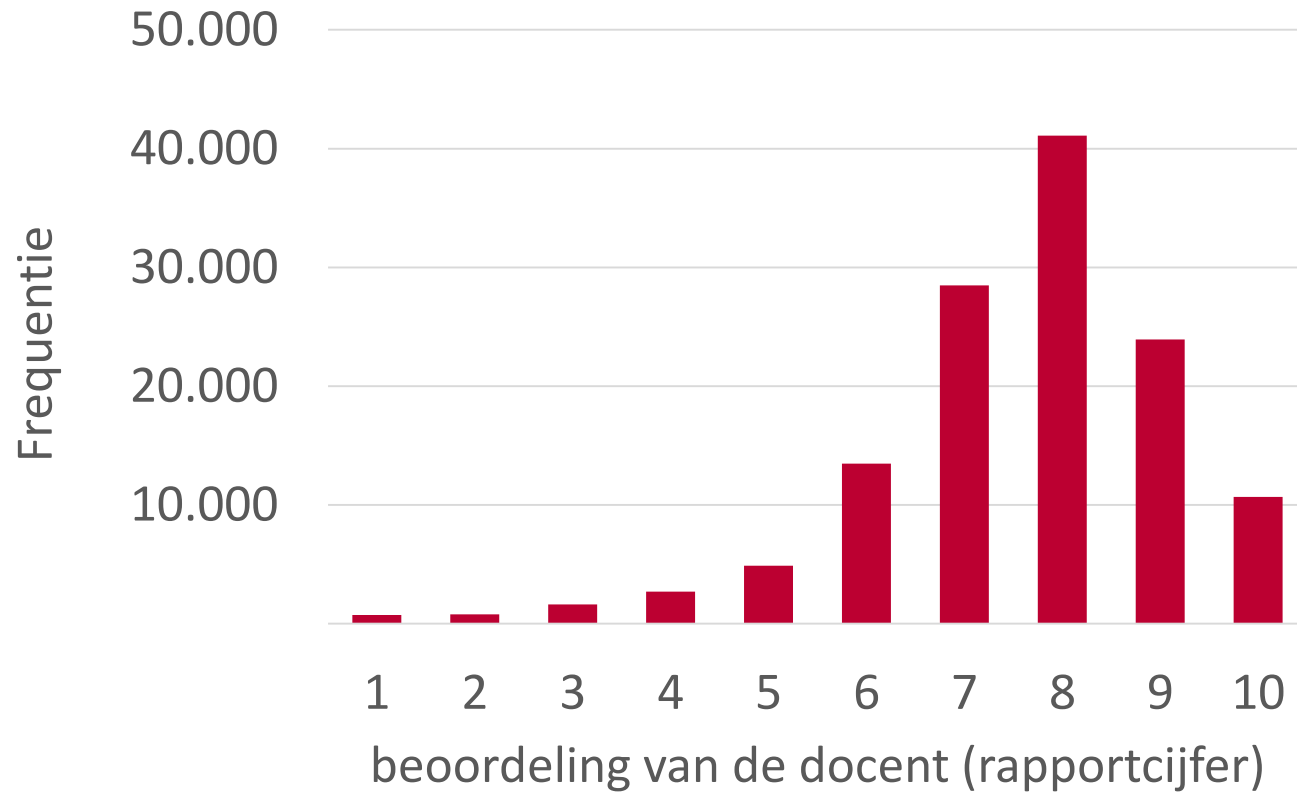
## Vakevaluaties:

- Verplichte vakken 2016/'17
- 1677 vakken/cursussen
- 4992 groepen/klassen
- 2234 verschillende docenten
- 5 faculteiten
- 128.377 studentbeoordelingen van hun docenten



# Data

(respons)



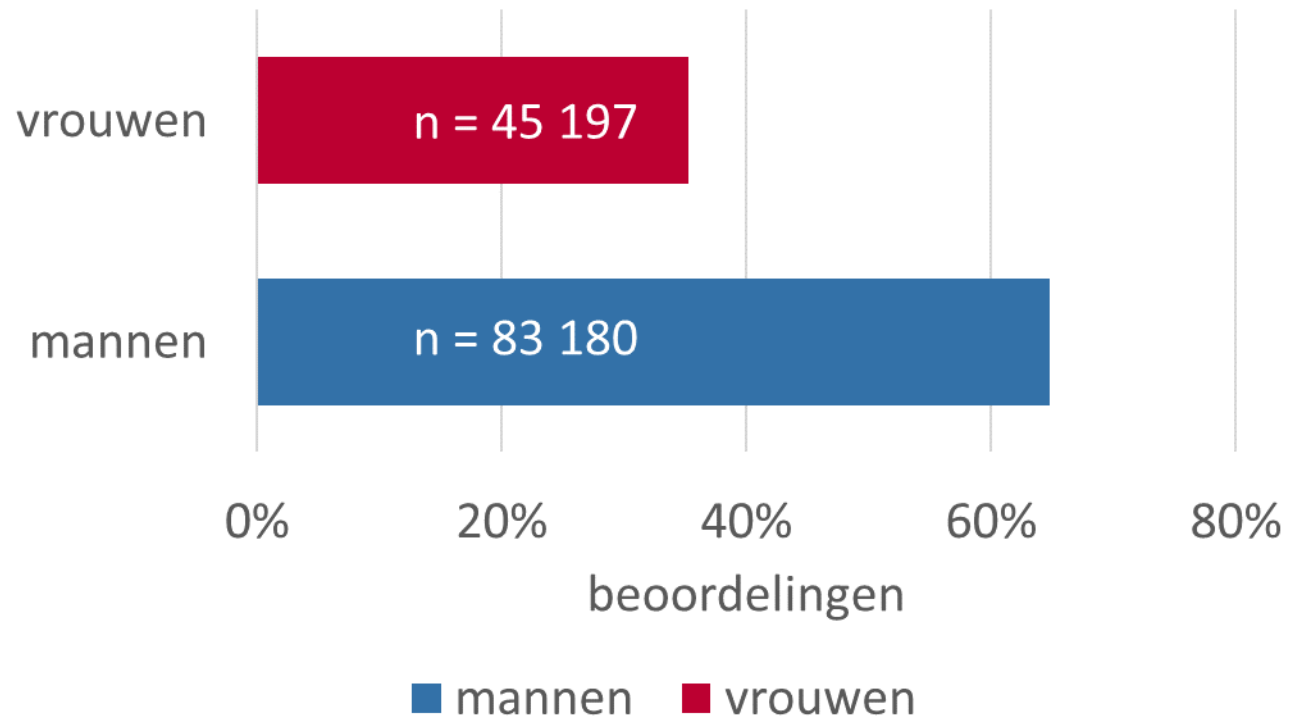
# Data

Achtergrondkenmerken:

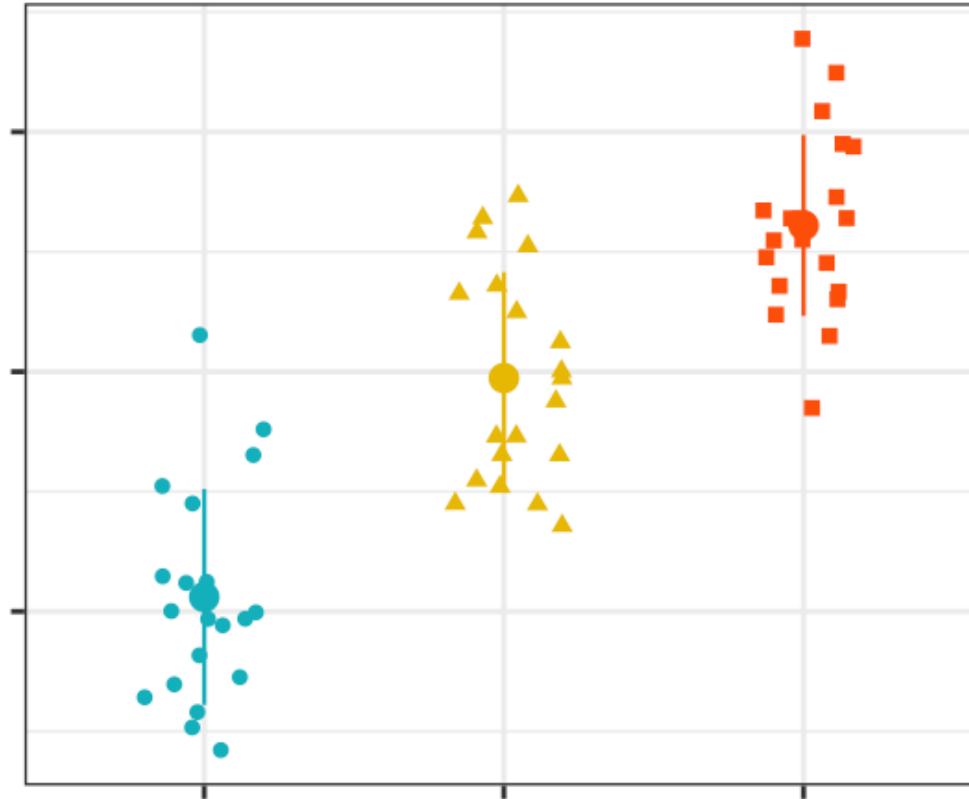
- **docent:** geslacht, nationaliteit (NL/niet NL), leeftijd, dienstverband (vast/tijdelijk), deeltijdfactor, functie, dienstjaren bij de UvA, onderwijskwalificatie
- **cursus:** faculteit, cursusniveau (ba/ma), groepsgrootte vak,
- **survey:** online/papieren survey, evaluatietaal



# Data



# Methode



ICC = 0,27



# Methode

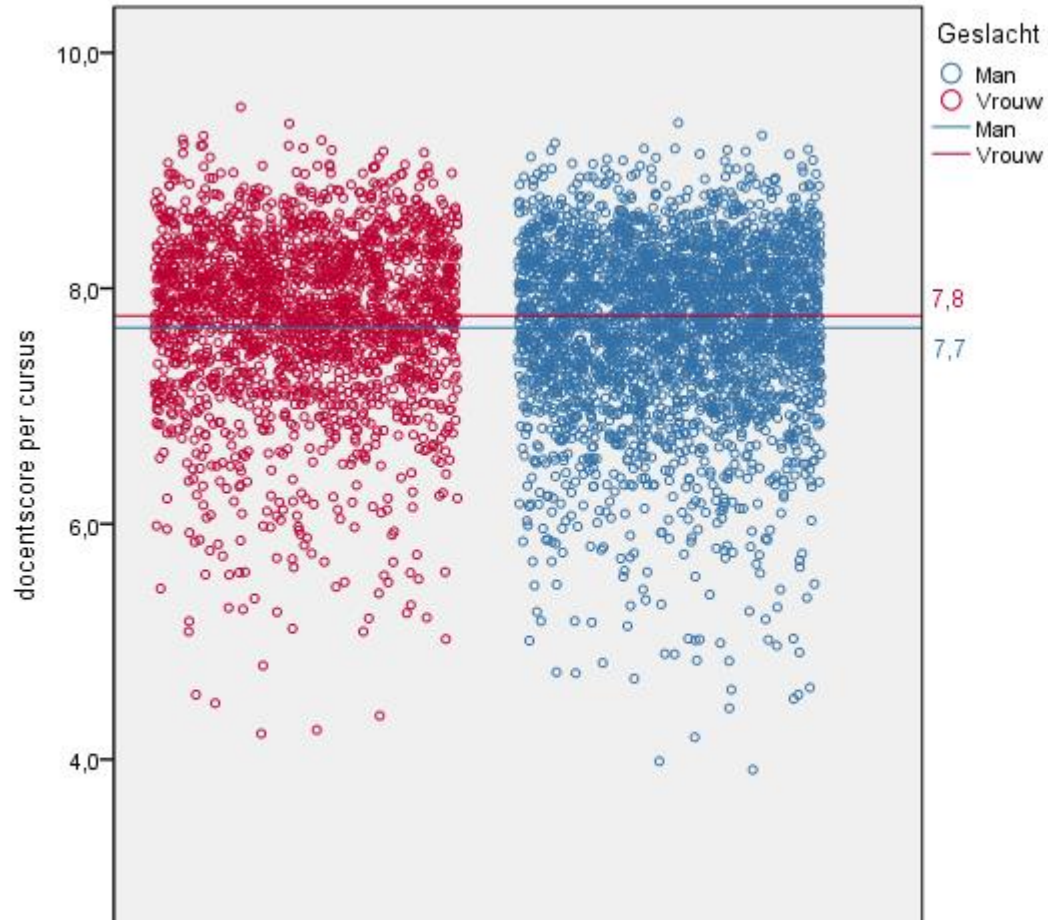
Modellen:

- 1) alleen sekse om docentbeoordeling per sekse vast te stellen (i.p.v. onafhankelijke  $t$ -toets)
- 2) volledig model om samenhang achtergrondvariabelen en docentscores te onderzoeken



# Resultaten

# Docentscores naar geslacht





# Samenhang docentscores en achtergrondvariabelen

## **Samenhang in het model:**

Leeftijd en senioriteit

Klassengrootte

Type college (werk-/hoorcollege)

Functie

Faculteit

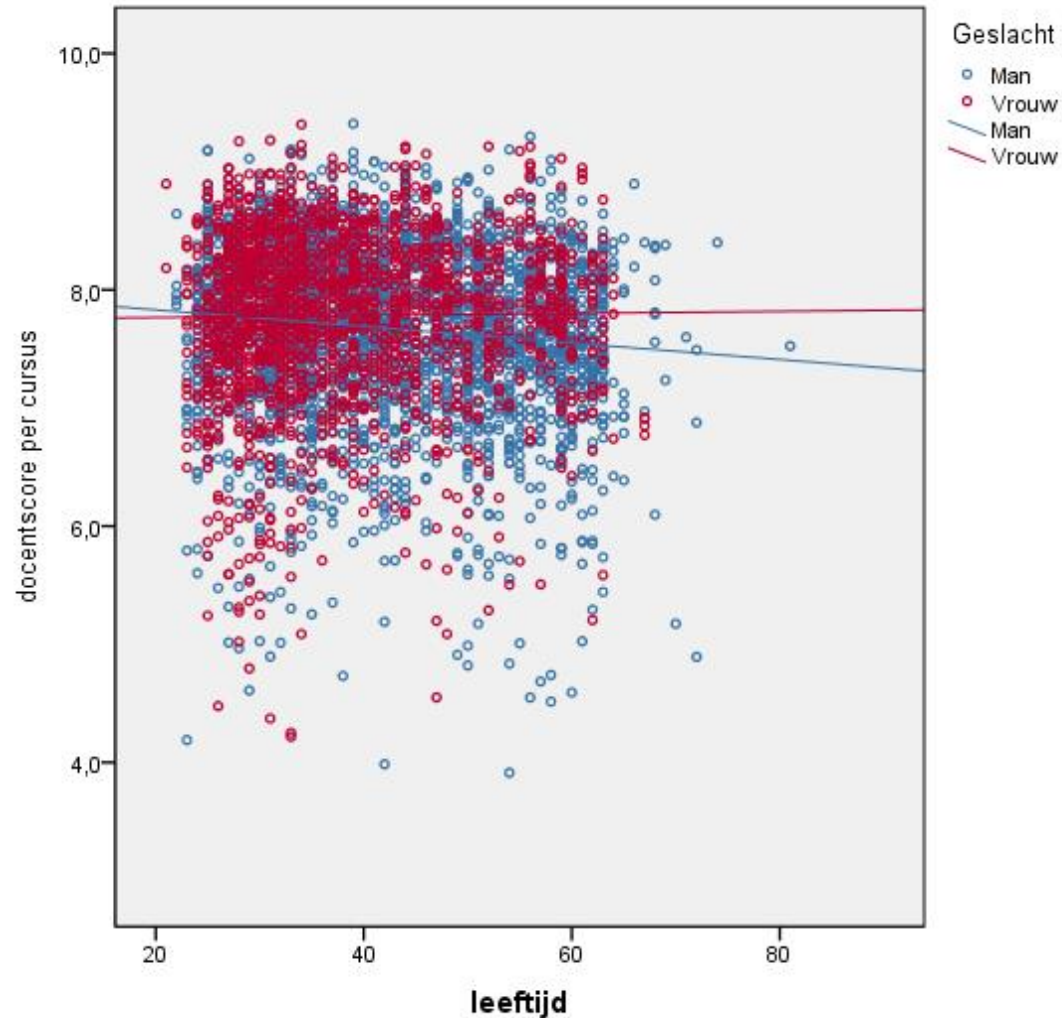
Master/bachelor

Online survey /papieren survey

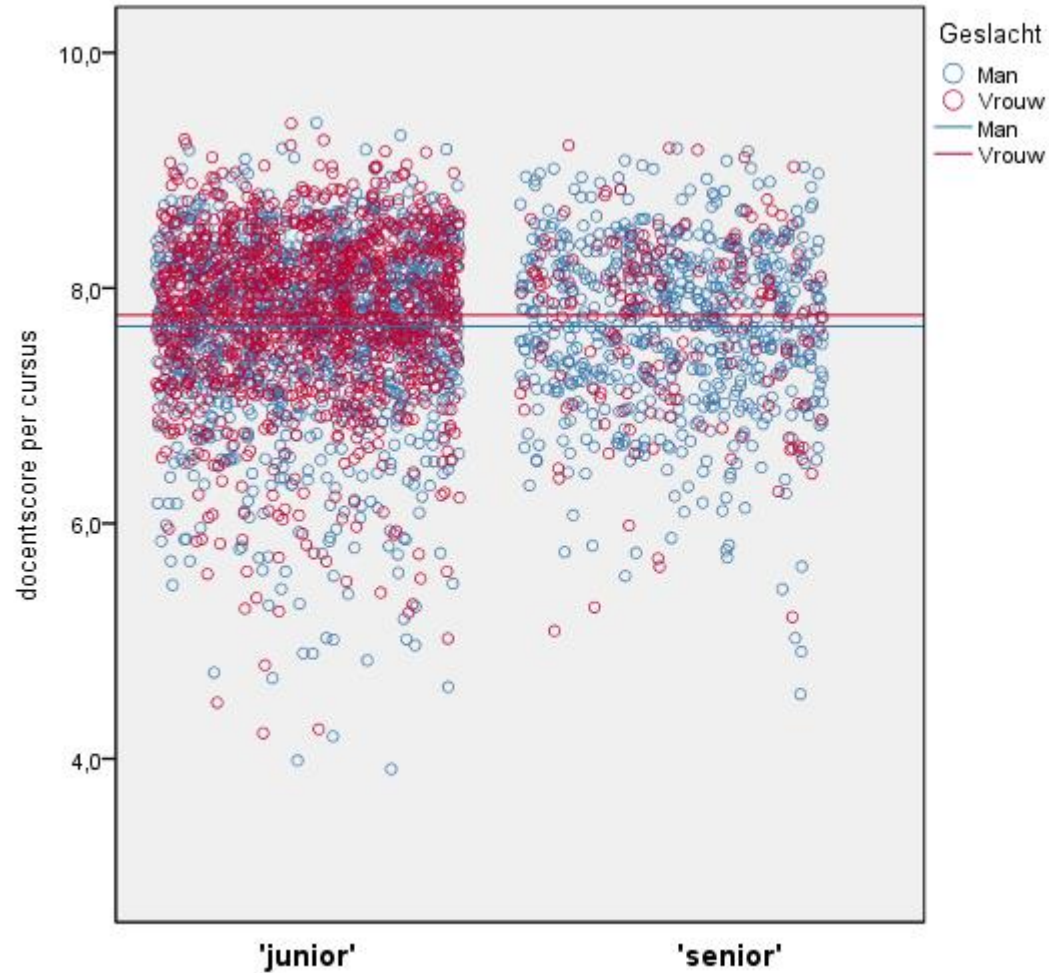
## **Geen samenhang in het model:**

Geslacht, nationaliteit, onderwijskwalificatie, vaste/tijdelijke aanstelling, taal van de survey, deeltijdfactor

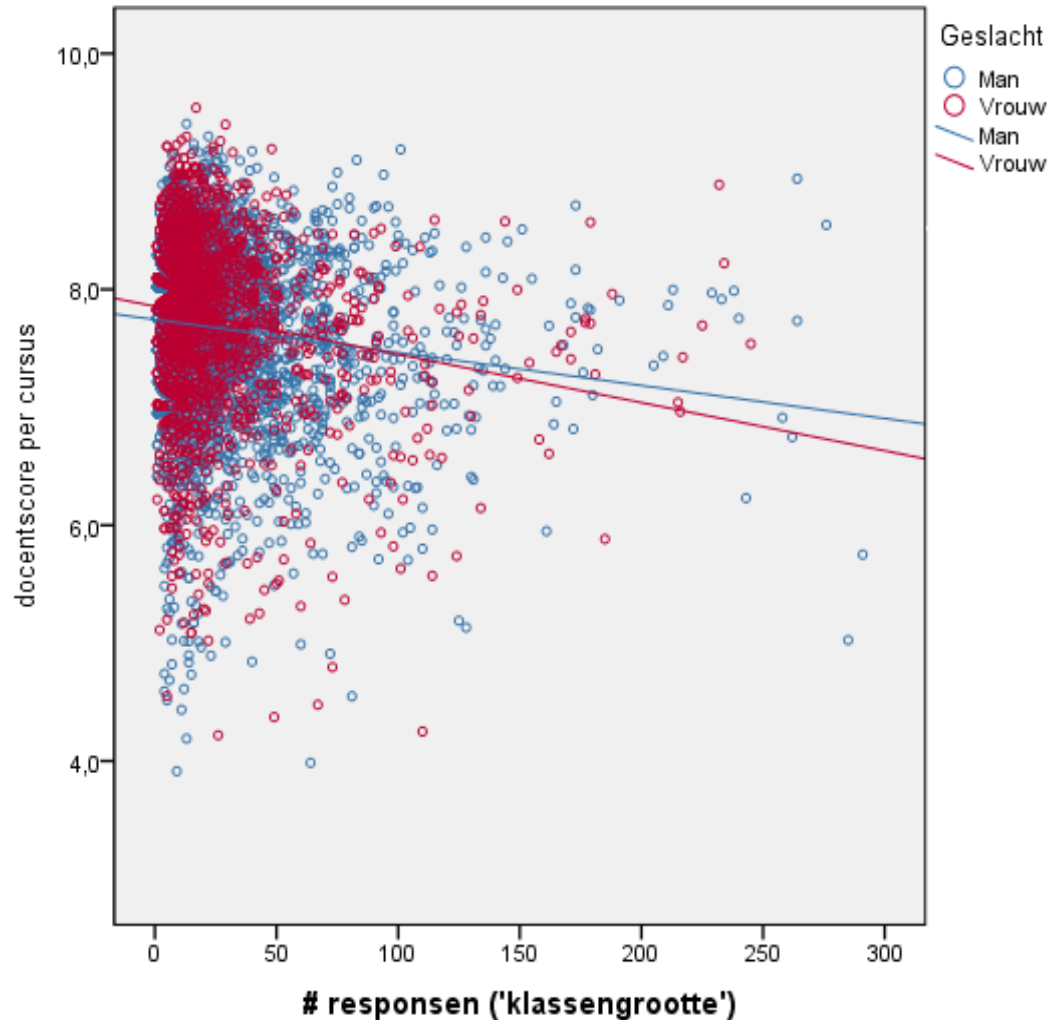
# Docentscores naar leeftijd



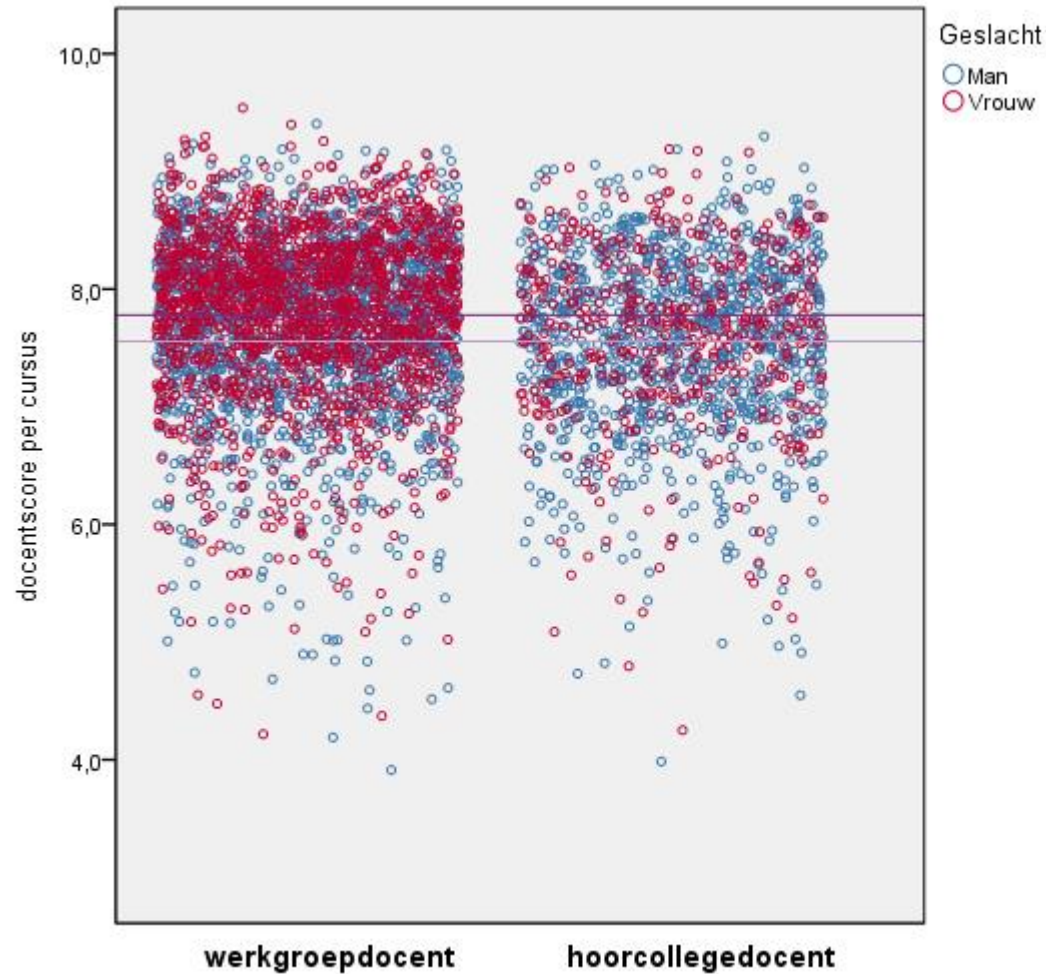
# Docentscores naar functie



# Docentscores naar klassengrootte



# Docentscores naar soort gegeven college



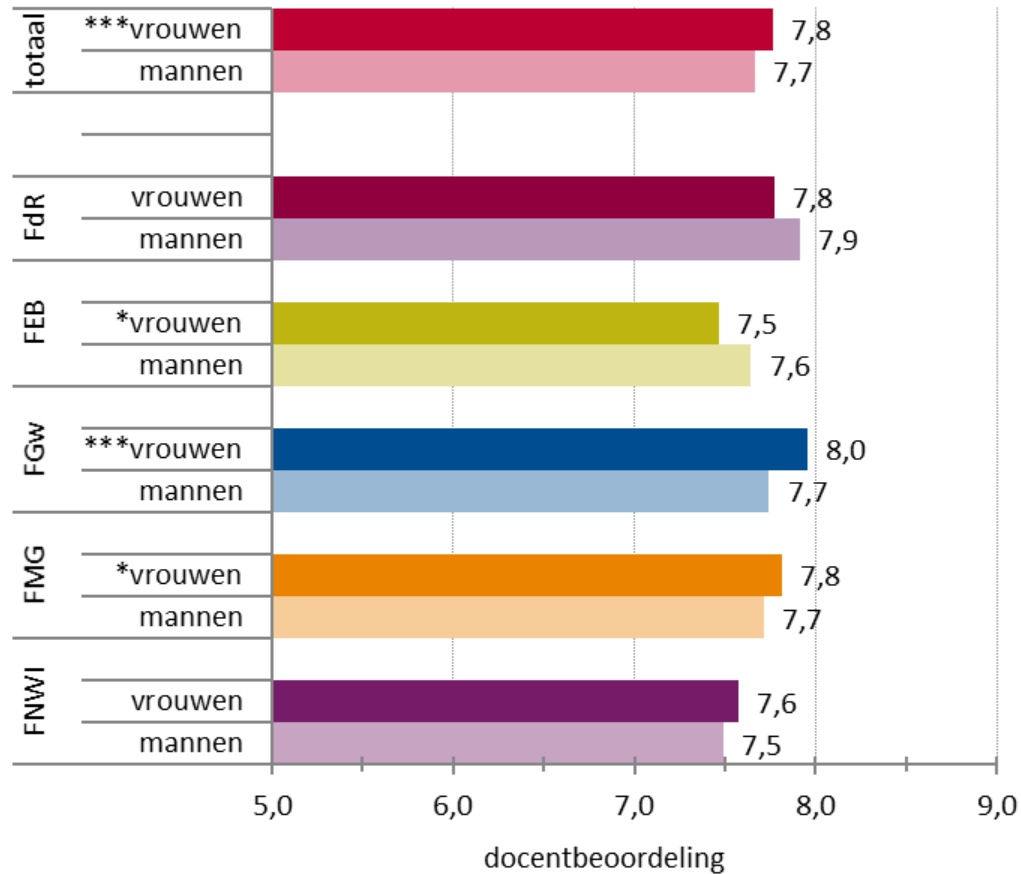
# Deelaspecten van het doceren

Differences in particular aspects of lecturing, schematic summary of gender only-model results

| Type of class             | item                                                                     |
|---------------------------|--------------------------------------------------------------------------|
| Lecture                   | The lecturer provided good examples to clarify the subject matter        |
| Lecture                   | The lecturer brought the subject matter to life                          |
| Lecture                   | The lecturer demonstrated the importance of the subject matter           |
| Lecture                   | The lecturer encouraged the students to think about the subject matter   |
| <b>Lecture</b>            | <b>The lecturer explained the subject matter clearly<sup>1)</sup></b>    |
| <b>Seminar / tutorial</b> | <b>I was given clear instructions for the assignments and activities</b> |
| <b>Seminar / tutorial</b> | <b>The lecturer demonstrated why the assignments were useful</b>         |
| <b>Seminar / tutorial</b> | <b>The supervision during the sessions was good</b>                      |
| <b>Seminar / tutorial</b> | <b>The lecturer provided good feedback</b>                               |
| Seminar / tutorial        | The lecturer encouraged the students to think about the subject matter   |
| <b>Seminar / tutorial</b> | <b>The lecturer encouraged the students to participate actively</b>      |
| <b>Seminar / tutorial</b> | <b>The sessions were well organised</b>                                  |

1) Marked in bold are the items for which a significant difference between male and female lecturers was predicted within the gender only-models. In the full models, no significant differences between male and female lecturers were found.

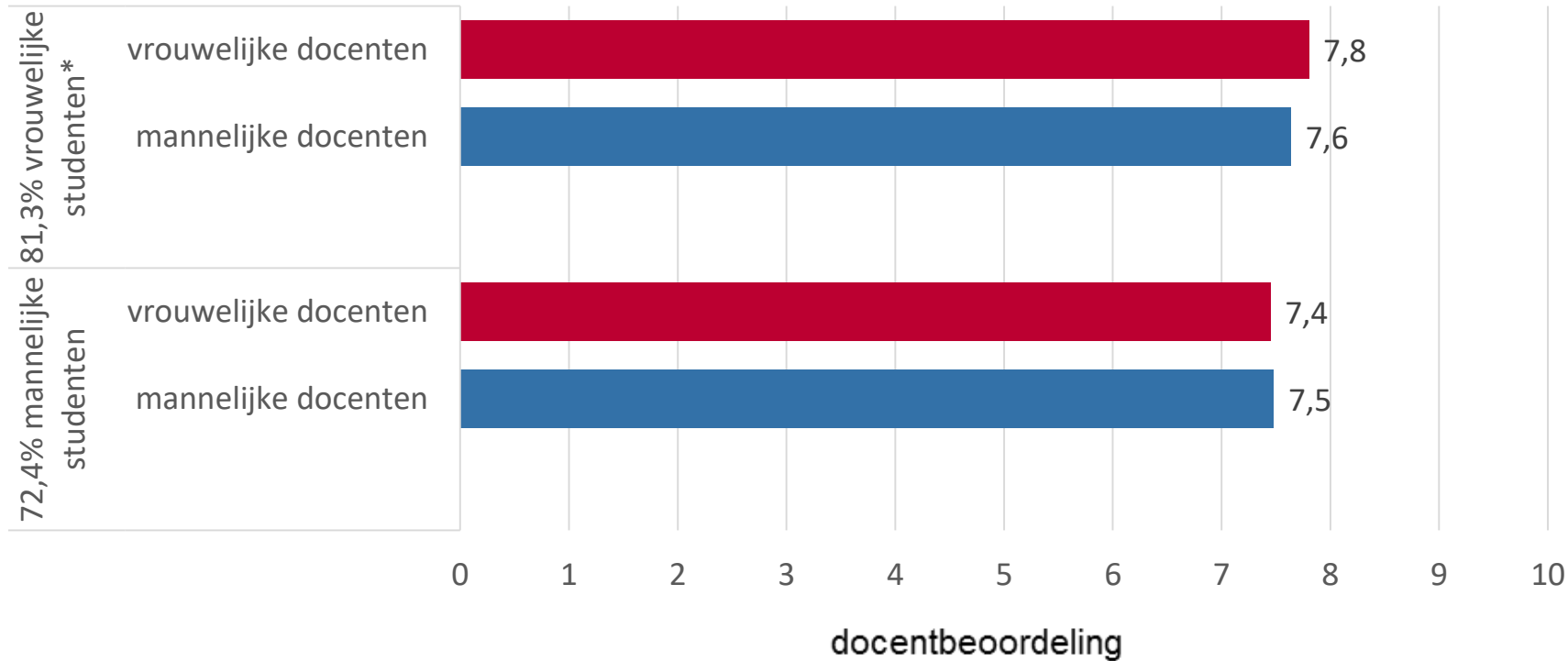
# Overige resultaten: faculteiten



\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$

# Overige resultaten: sekse studenten

Top-10 opleidingen met grootste aandeel mannelijke / vrouwelijke studenten



\*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$

# Conclusies

Vrouwelijke docenten worden hoger beoordeeld dan mannelijke docenten als we naar de UvA als geheel kijken. De verschillen zijn echter klein en de variatie binnen de groepen is groot.

De data suggereren dat de verschillen samenhangen met verschillen in 'demografische kenmerken' van de groepen vrouwelijke en mannelijke docenten.

Binnen de Faculteit der geesteswetenschappen waren de sekseverschillen in docentscores het grootst.

# Wat hebben met de resultaten gedaan?

## Fact

19 juli 2018  
International Research & Culture week

### SEKSEVERSCHILLEN IN DOCENTBEOORDELING

Studentbeoordelingen van mannelijke en vrouwelijke docenten in UvA

- Vrouwelijke docenten krijgen gemiddeld een net wat hoger rapportcijfer dan mannelijke docenten.
- Vrouwelijke docenten bij de FMG en FGW krijgen een hoger rapportcijfer en bij de FEB een lager rapportcijfer. Bij de PWW en FdR zijn er geen significante verschillen.
- Vrouwelijke docenten scoren op bijna alle facetten van het doceren waarvoor gevraagd wordt hoger dan mannelijke docenten.

Van tijd tot tijd komt de vraag op of er niet een genderbias zit in de manier waarop studenten hun docenten beoordelen in de vakbeoordelingen. De wetenschappelijke literatuur gaat op dit punt geen eenduidig antwoord. Empirisch onderzoek wijst in verschillende richtingen. Reden genoeg om eens met behulp van UvAdata te onderzoeken welke verschillen er boven water te halen zijn in de UvA Q vakbeoordelingen, waarin studenten onder meer de vraag krijgen hun docenten een rapportcijfer (1 tot en met 10) te geven.

#### Resultaten

Uit de analyses blijkt dat studenten vrouwelijke docenten gemiddeld 0,1 punten ( $p < 0.001$ ) hoger beoordeelden dan mannelijke docenten die gemiddeld een 7,7 scoren.

Uitgesplitst naar faculteit zijn de docentscores van vakken bij de FGW en FMG hoger voor vrouwen dan voor mannen (resp. 0,2 punten,  $p < 0.001$  en 0,4 punten,  $p < 0.05$ ). Bij de FEB scoren vrouwen juist lager (0,1 punten;  $p < 0.05$ ). Bij de FdR en PWW zijn er geen significante verschillen.



Uit de analyses blijkt docentbeoordelingen:

- het type docent: w/werklcollega/docent hoorcollege/docent
- de senioriteit: do dienst bij de UvA is
- de faculteit: door krijgen lagere score faculteiten
- het type survey: de docentscores op de surveys
- de groepsgroottes: de groepsgroottes, hoe

UNIVERSITEIT VAN AMSTERDAM

## Fact

19 juli 2018  
International Research & Culture week

### GENDER DIFFERENCES IN TEACHING EVALUATIONS

Assessment scores of male and female lecturers in UvA Q

- The assessment scores of female lecturers are slightly higher (+0.1 points on a scale from 1 to 10;  $p < 0.001$ ) than those of male lecturers.
- Female lecturers at the Faculty of Humanities ( $p < .001$ ), and the Faculties of Economics and Business, and Social and Behavioural Sciences ( $p < .05$ ) receive higher assessment scores than their male colleagues ( $p < .001$ ). At the Faculty of Science and Law there are no significant differences.
- Female lecturers score better than male lecturers on most UvA Q-questions about aspects of teaching.

Every now and then the question arises whether there is a gender bias in the way students assess their lecturers in course evaluation surveys. This may question the legitimacy of course evaluation surveys. The academic literature does not offer a clear conclusion. Empirical research has shown different outcomes. Reason enough to examine with UvAdata which gender differences can be found in the UvA Q course evaluations.

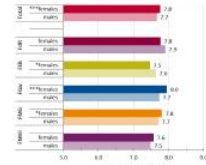
#### Results

Multilevel analysis shows that the assessment scores of female lecturers are slightly higher (+0.1 points,  $p < .001$ ) on a scale from 1 to 10 than those of male lecturers.

Broken down by faculty, the results show that female lecturers at the Faculties of Humanities and Social and Behavioural Sciences score significantly higher than male lecturers (resp.

+0.2 points,  $p < .001$ ; +0.1 points,  $p < .05$ ). At the Faculty of Economics and Business female lecturers receive lower assessment scores than their male colleagues (-0.1 points,  $p < .05$ ). There are no differences between the assessment scores at the Faculties of Law and Science.

Figure 3: Average assessment scores (scale 1 to 10)



Furthermore, analysis of more extensive models shows that the assessment scores of lecturers are associated with:

- Type of lecturing and class size: Lecturers who facilitate tutorials and seminars receive (on average) higher assessment scores than those who facilitate lectures.
- Furthermore, the smaller the class size, the higher the assessment score of the lecturer tends to be.
- Age and seniority: The older the lecturer, the lower his/her assessment score. The effect is stronger for male lecturers than for female lecturers. Furthermore, a seniority of 15 years

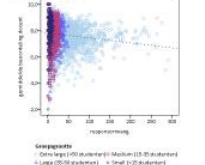
Figure 2: Samenhang responsiviteit, groepsgroottes en docentbeoordeling

### Wijzen de (gecorrigeerde) sekseverschillen op een genderbias?

Een verschil tussen mannelijke en vrouwelijke

- Faculty: On average, lecturers at the Faculty of Economics and Business and the Faculty of Sciences receive lower assessment scores.
- Type of survey: On average, the assessment scores for lecturers are lower when online surveys and not paper surveys are used (-0.1 points,  $p < .001$ ).

Figure 2: Relation between response size, class size and assessment scores



These and other background characteristics (see section 'Analysis') are not evenly distributed over the groups of male and female lecturers. For instance, there are relatively more male lecturers who are more than 15 years in the service of the UvA, and the class sizes of courses taught by female lecturers are on average smaller. Controlled for these differences in background characteristics, there are no significant differences between the assessment scores of male and female lecturers at the UvA.

On the level of the faculties, and after controlling for the background variables, there are no significant differences in assessment scores between the Faculties of Economics and Business, Social and Behavioural Science and the Faculty of Science. At the Faculty of Law, male lecturers score 0.2 points ( $p < .05$ ) higher and at the Faculty of Humanities 0.2 points ( $p < .01$ ) lower.

#### Relation background characteristics and gender

Looking at the association between background variables and the assessment scores for male and

female lecturers separately, the following can be noticed:

Male PhD's receive lower assessment scores in comparison to male assistant professors (-0.3 points;  $p < .01$ ). For female lecturers there is no difference in assessment scores between PhD's and assistant professors (UD's).

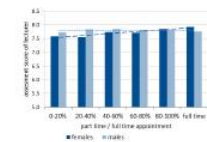
#### Do gender differences imply gender bias?

The differences found between male and female lecturers do not necessarily imply a gender bias. Nor does the absence of differences point at the absence of a bias. In practice it is impossible to control for all variables other than gender. Also, some of the gender differences measured might be caused by valid measurements of true difference. Therefore, it is practically impossible to conclude decisively if, and to what extent, a gender bias is at play.

Although it is practically impossible to confirm or to deny the existence of a gender bias with this type of study, these exploratory findings of gender differences in teaching evaluations have given a good impression of 1) the score of the gender differences and 2) the differences in the way various background characteristics are associated with the assessment scores of both male and female lecturers.

The assessment scores of female professors are an estimated 0.4 points lower ( $p < .001$ ) than those of female assistant professors. For male lecturers there is no significant difference between the assessment scores of professors and assistant professors.

Figure 3: Assessment scores and part-time appointment



The assessment scores of female lecturers with a part-time appointment are lower than the scores



# Dank voor je aandacht !

## Vragen?

Meer weten? Mail naar [j.j.a.m.vandentillaart@uva.nl](mailto:j.j.a.m.vandentillaart@uva.nl)