Gender and Performance in Collaborative Research: Evidence From Student Teams

Max Coveney Teresa Bago d'Uva Pilar García-Gómez

Applied Economics Department
Erasmus University Rotterdam

March 7, 2022

Introduction

- CLI project: *Capturing and realising the benefits of diversity at Erasmus University*
- Goal 1: implement diversity policy for 1st year bachelor students
- Goal 2: use policy to investigate impact of (gender) diversity on student group performance

Diversity policy

- Undergraduate Economics program at Erasmus (Dutch & English)
- $\sim 1,300$ 1st year student across two cohorts (2018-19 & 2019-20)
- Students participate in 3-block long course with focus on academic/research tasks, done in research teams (pairs)
  - Block 3 – Writing (synthesizing literature, motivating etc.)
  - Block 4 – Data (collection, wrangling, analyzing etc.)
  - Block 5 – Research paper + Presentation
- Pre-policy: students clustering by gender, ethnicity, nationality etc
- Our policy involves randomly assigning students teams to promote diversity (contact hypothesis)

Goal 2

Research question: how does the gender composition of student research teams affect their performance (in terms of grades)?

Two motivations:

- Educational implications: how should we form student teams to improve learning outcomes?

- Research implications: can we generalize these student teams to other (research) teams?

Research implications

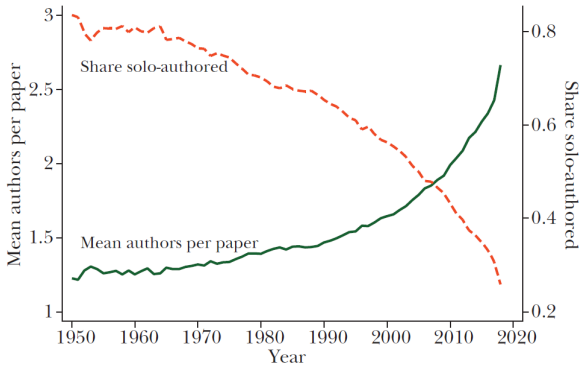
Teams increasingly important in research occupations

- Majority of research papers in Science and Engineering and Social Sciences, and majority of filed US patents, now written in teams (Wuchty et al., 2007)

Research implications

Teams increasingly important in research occupations

- Majority of research papers in Science and Engineering and Social Sciences, and majority of filed US patents, now written in teams (Wuchty et al., 2007)



Research implications

Observational data suggests *gender* composition of team important:

- Yang et al. (2021) show that gender diverse teams produce more “innovative” and more cited work in medical science
- Hengel (2021) and Hengel & Moon (2021) show that Economics papers with more female authors are better written and are cited more

This presentation

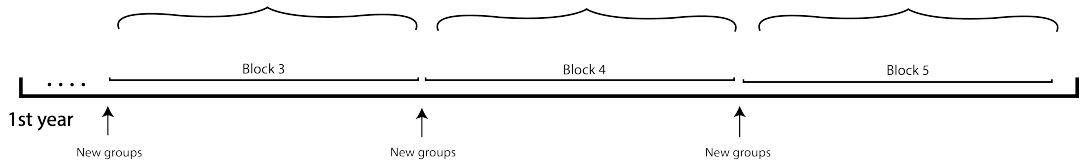
- Students perform graded “research-like” tasks in randomly allocated teams
- Gender composition important for performance: all-males pairs outperformed by other combinations
- Effect survives comprehensive “ability” controls for each member of pair

Course structure

- Task 1 (10%) – Find, assess and summarize academic articles
- Task 2 (20%) – Write introduction & main body of lit. review
- Task 3 (10%) – Give feedback on assignment 2
- Task 4 (60%) – Write final literature review

- Task 5 (20%) – Collect data from existing databases
- Task 6 (10%) – Set up and distribute own survey
- Task 7 (20%) – Data analysis
- Task 8 (50%) – Write up data analysis

- Task 9 (20%) – Write research proposal
- Task 10 (10%) – Give feedback on proposal
- Task 11 (30%) – Write research paper
- Task 12 (40%) – Presentation



Research “tasks”

- Writing
  - Summarize existing articles
  - Write-up empirical results
  - Write research paper
- Data
  - Collect existing data
  - Run survey
  - Analyses data
- Presentation
  - Prepare & give presentation on research paper
- Feedback
  - Evaluate other assignments
  - Provide feedback

Data

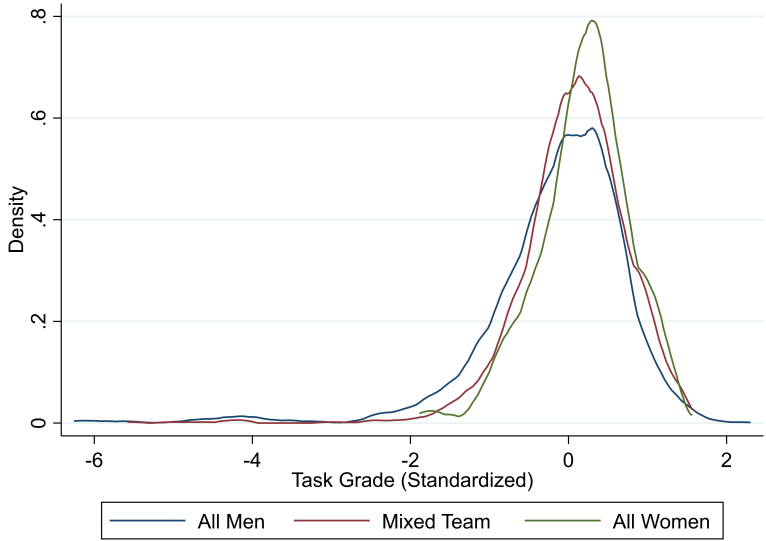
- Diversity policy data
  - Assigned groups
  - Task performance (grades)
  - Tutorial group
- Administrative university data
  - Age, gender, ethnicity, nationality
  - Parents' education level $\Rightarrow$ SES measure
  - High school GPA (Dutch students)
  - All course results $\Rightarrow$ University GPA

Data

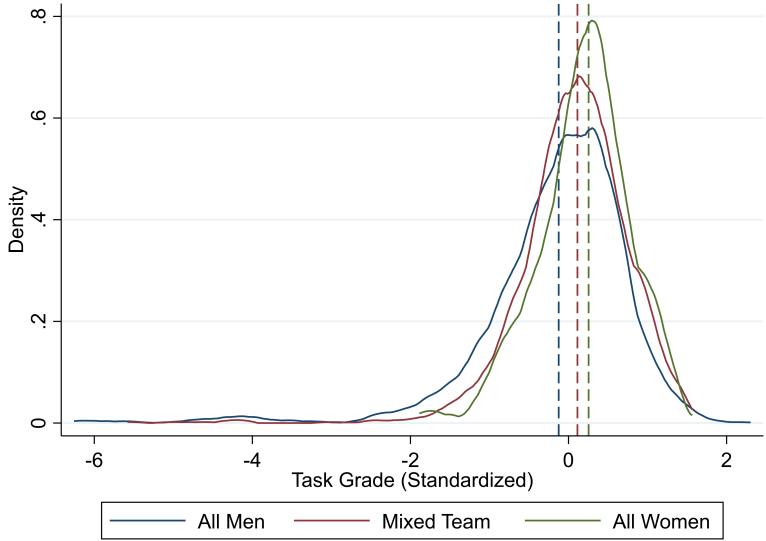
	Mean	SD	Count
Student Data			
Number of students			1,281
Number of blocks present	2.069	(0.776)	1,281
Student is female	0.300		1,281
Age on October 1st	18.528	(1.104)	1,281
GPA before block 3 in first year	6.691	(0.996)	1,281
High school GPA	7.001	(0.652)	987
Non-Dutch	0.169		1,279
Native Dutch	0.617		1,279
Immigrant Dutch (West)	0.062		1,279
Immigrant Dutch (Non-West)	0.152		1,279
Both parents attended university	0.346		1,027
Group Data			
Number of groups			1,053
Number of groups in Block 3			478
Number of groups in Block 4			201
Number of groups in Block 5			374
All men	0.493		1,053
Gender mix	0.416		1,053
All women	0.091		1,053

	Mean	SD	Count
Task Data			
Average block grade	73.484	(9.262)	1,053
Average block grade Block 3	74.590	(8.515)	478
Average block grade Block 4	70.027	(11.109)	201
Average block grade Block 5	73.928	(8.650)	374
Average task grade	72.778	(13.261)	4,212
Average task grade Writing	71.681	(13.015)	2,383
Average task grade Data	67.794	(13.108)	603
Average task grade Presentation	74.485	(9.197)	374
Average task grade Feedback	78.626	(13.454)	852

Task results



Task results



Empirical approach

$$Grade_{trg} = \beta_0 + \beta_1 Mixed_r + \beta_2 AllWomen_r + Task_t + Tut_g + \epsilon_{trg}$$

- $Task_t$ - Assignment fixed effects
- Tut_g - Tutorial group fixed effects

Empirical approach

$$Grade_{trg} = \beta_0 + \beta_1 Mixed_r + \beta_2 AllWomen_r + Task_t + Tut_g + \sum_{q=1}^{4} \theta_{1q} \mathbb{1}\left(AbilityQuintile_r^{Best} = q\right) + \sum_{p=1}^{4} \theta_{2p} \mathbb{1}\left(AbilityQuintile_r^{Worst} = p\right) + \epsilon_{trg}$$

- $AbilityQuintile_r^{Best}$ - Ability quintiles for best in team
- $AbilityQuintile_r^{Worst}$ - Ability quintiles for worst in team
- Ability controls: High school GPA/University GPA

How good are ability controls?

Regression results

	(1)	(2)	(3)
	Task Grades (Std)		
Mixed Team	0.222***	0.244***	0.214***
	(0.0457)	(0.0450)	(0.0481)
All Women	0.319***	0.359***	0.305***
	(0.0694)	(0.0718)	(0.0679)
Best/Worst GPA Quint. (Uni)		✓	
Best/Worst GPA Quint. (HS)			✓
Mixed Team=All Women			
F-Statistic	3.180	3.840	2.890
p-value	0.077	0.052	0.092
Observations	4,212	4,212	3,744

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Further results

Extensions & robustness checks:

- Per-task type analysis
- Results by performance percentile
- Non-gender characteristics analysis
- Teams $>$ 2 analysis

Marker bias | Alternative ability controls

Per task type

	Writing			Data		
	(1)	(2)	(3)	(4)	(5)	(6)
	Task Grades (Std)					
Mixed Team	0.222***	0.238***	0.203***	0.304***	0.358***	0.294***
	(0.0542)	(0.0517)	(0.0582)	(0.0942)	(0.100)	(0.106)
All Women	0.274***	0.317***	0.267***	0.362***	0.397***	0.252**
	(0.0832)	(0.0855)	(0.0776)	(0.114)	(0.121)	(0.114)
Best/Worst GPA Quint. (Uni)		✓			✓	
Best/Worst GPA Quint. (HS)			✓			✓
Mixed Team=All Women						
F-Statistic	0.636	1.203	1.072	0.273	0.158	0.214
p-value	0.427	0.275	0.303	0.603	0.693	0.646
Observations	2,383	2,383	2,117	603	603	534

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Per task type

	Presentation			Feedback		
	(1)	(2)	(3)	(4)	(5)	(6)
	Task Grades (Std)					
Mixed Team	0.276***	0.294***	0.339***	0.163**	0.170**	0.151*
	(0.0967)	(0.101)	(0.0961)	(0.0728)	(0.0747)	(0.0814)
All Women	0.282*	0.243	0.350*	0.325***	0.334***	0.302***
	(0.167)	(0.165)	(0.210)	(0.109)	(0.108)	(0.100)
Best/Worst GPA Quint. (Uni)		✓			✓	
Best/Worst GPA Quint. (HS)			✓			✓
Mixed Team=All Women						
F-Statistic	0.001	0.110	0.003	2.690	2.660	2.540
p-value	0.974	0.741	0.955	0.103	0.105	0.114
Observations	374	374	335	852	852	758

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

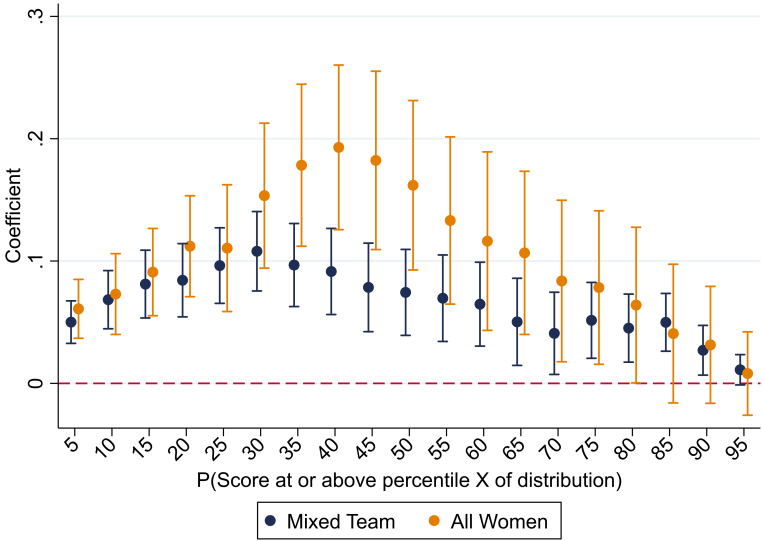
Further results

Extensions & robustness checks:

- Per-task type analysis
- Results by performance percentile
- Non-gender characteristics analysis
- Teams $>$ 2 analysis

Marker bias | Alternative ability controls

Per performance percentile



Further results

Extensions & robustness checks:

- Per-task type analysis
- Results by performance percentile
- Non-gender characteristics analysis
- Teams $>$ 2 analysis

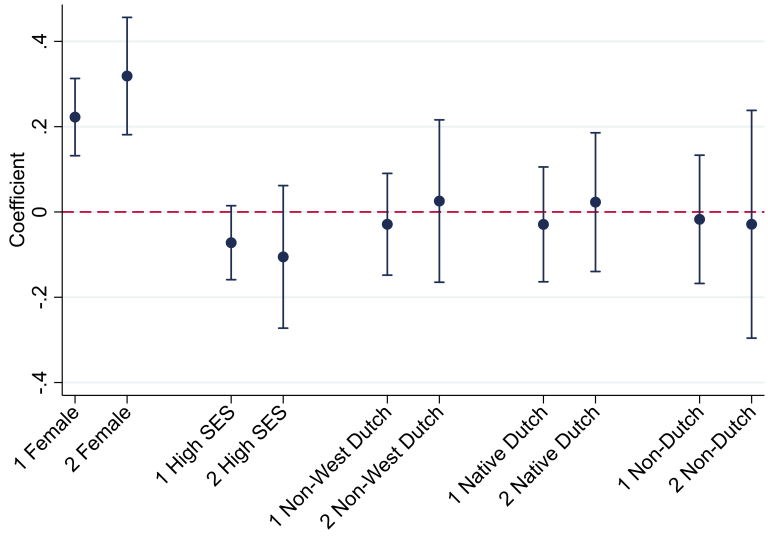
Marker bias | Alternative ability controls

Non-gender characteristics

	(1)	(2)	(3)
	Task Grades (Std)		
Mixed Team	0.244***	0.224***	0.231***
	(0.0450)	(0.0496)	(0.0507)
All Women	0.359***	0.320***	0.330***
	(0.0718)	(0.0729)	(0.0746)
Best/Worst GPA Quint. (Uni)	✓	✓	✓
Dutch Ethnicity Controls		✓	✓
Dutch Nationality Controls		✓	✓
SES Controls			✓
Mixed Team=All Women			
F Statistic	3.840	3.570	2.680
p-value	0.052	0.061	0.104
Observations	4,212	4,212	3,744

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Non-gender characteristics



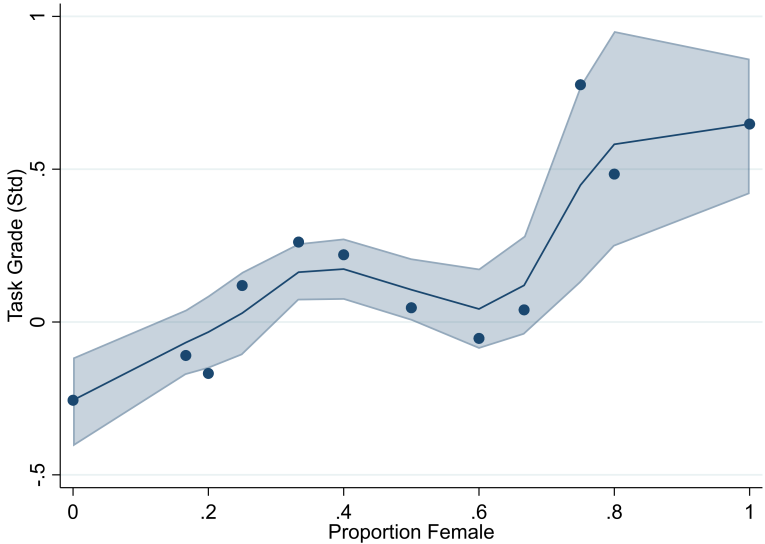
Further results

Extensions & robustness checks:

- Per-task type analysis
- Results by performance percentile
- Non-gender characteristics analysis
- Teams $>$ 2 analysis

Marker bias | Alternative ability controls

Larger teams analysis



Larger teams analysis

	(1)	(2)	(3)	(4)
	Task Grades (Std)			
Mixed Team	0.258	0.263*	0.299*	0.297*
	(0.167)	(0.146)	(0.151)	(0.150)
All Women	0.428**	0.335*	0.304	0.342
	(0.201)	(0.180)	(0.231)	(0.243)
Avg. University GPA		✓		✓
Best/Worst University GPA Quint.			✓	✓
Mixed Team=All Women				
F-Statistic	3.15	0.59	0.001	0.0481
p-value	0.0817	0.446	0.979	0.827
Observations	604	604	604	604

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Wrapping up

Results so far:

- Gender composition of student research team matters
- Effect not due to differences in individual ability between men/women
- Gender difference in “group work skills”?

Wrapping up

Results so far:

- Gender composition of student research team matters
- Effect not due to differences in individual ability between men/women
- Gender difference in “group work skills”?

Implications:

- Direct evidence for “quality” difference found in endogenous teams (Hengel, 2021; Hengel & Moon, 2021)
- Findings may have implications for optimal team formation in research settings
- “Leaky pipeline” in economics - even more problematic?

Wrapping up

Next steps:

- Further data collection
- Teams in other contexts
- Team survey to understand mechanisms

Thank you for your attention!

coveney@ese.eur.nl
marreirosbagoduva@ese.eur.nl
garciagomez@ese.eur.nl

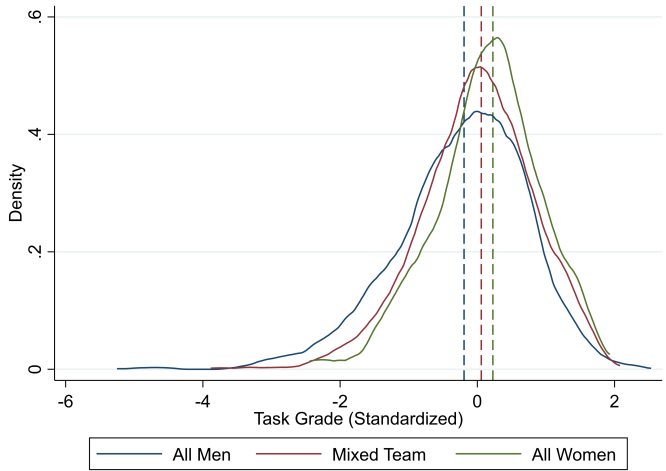
Individual Grade Results

	(1)	(2)	(3)
	Individual Course Results (Std)		
Female Student	0.0740*	0.0597	-0.00792
	(0.0446)	(0.0376)	(0.0546)
University GPA Quint.		✓	
Highschool GPA Quint.			✓
Observations	5,107	5,107	4,082

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Task Results No Zeros



Back

Example of Task (Task 2)

Writing the introduction of the literature review (± 300 words)
Use the introduction to introduce your subject. Make sure it includes the following:

- A (catchy) introduction of the topic.
- A good and well-explained research question.
- A clear explanation why the research is scientifically relevant. This means that you describe how your literature review adds to the existing academic literature.
- A clear explanation why the research is socially relevant. This means that you describe why it is important for society that research (here a literature review) has been performed on the specific topic.
- A description of the structure of the rest of your text.

The research question you have to answer in your literature review is:

'How does 'fill in your extension' affect the economic growth of countries?'

You can find more information on how to write a good introduction in the book 'Academic Writing Skills for Economics and Business Administration'.

Writing the main body of the literature review (± 1000 words)
In the book 'Academic Writing Skills for Economics and Business Administration' is described that the main body of an academic text consists of the following sections: theoretical framework, data & methodology, and results. Because you have to write a literature review, there is no datasource based on numbers, no model needs to be specified, and there also are no results based on numerical data. This is the reason that, in this assignment (and skills module), you only have to create the theoretical framework section to complete the main body of your literature review.

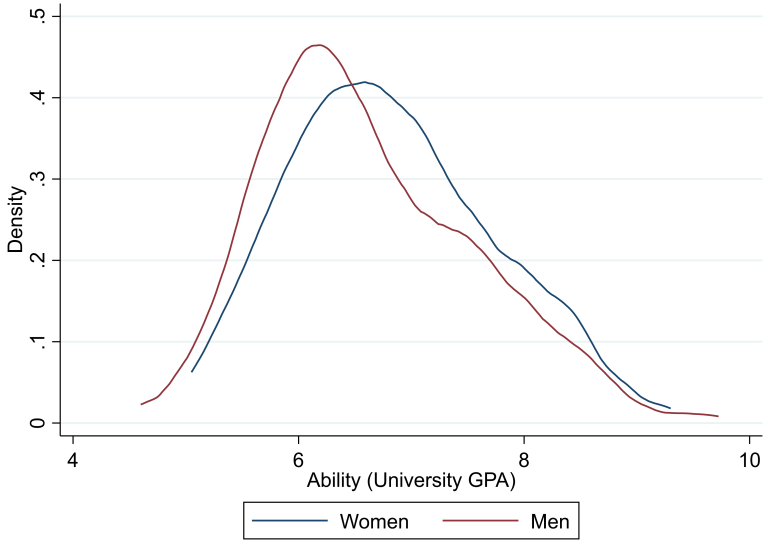
The main body is the largest section of your literature review. It is important that, in this section, you describe all the necessary information to answer your research question. It contains literature on the main topic as well as the extension. Altogether, the main body should include the following information:

- Definitions of the most important concepts of your review (including references). For example, think about the concepts economic growth and the extension you have chosen.
- A description of the essence of the 3 leading articles.
- A description of the essence of the articles about the extension (at least 2).

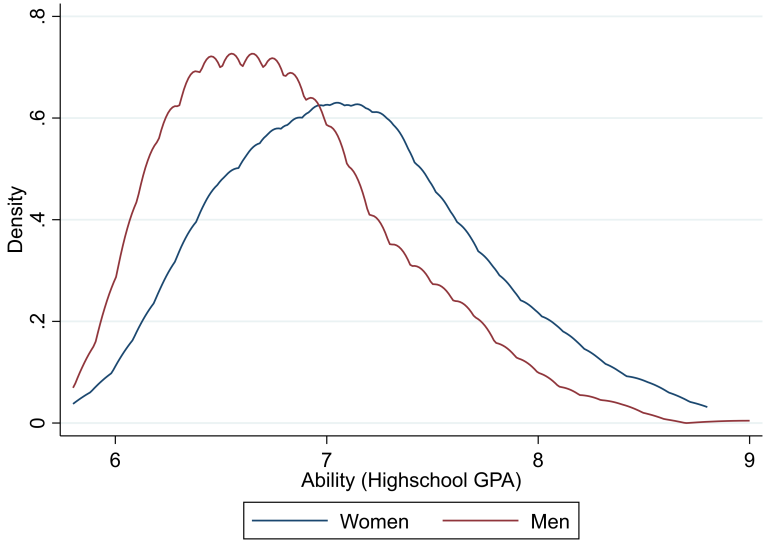
Assessment criteria		Max. points	Chapter[4]
The student has:			
Introduction			
1)	introduced the topic (in a catchy manner)	1	2
2)	formulated a good research question and explained it well	1	1, 2
3)	explained why the research is scientifically relevant	1	2, 7
4)	explained why the research is socially relevant	0.5	2,7
5)	described the structure of the rest of the text	0.5	2
Theoretical framework			
6)	defined the most important concepts	1	2, 3
7)	described the essence of the three leading articles	2	2, 3, 7
8)	described the essence of the articles about the extension	2	
9)	described the link between the articles used and the research question	3	1, 2, 3, 7
Sources			
10)	used a sufficient number of sources (3 given + 2 extension)	0,5	3, 6
11)	used relevant sources of good quality	0,5	3, 6
12)	referred correctly according to APA style (in the text)	1	6
13)	added a correct and complete bibliography according to APA style	1	6
Academic writing			
14)	written according to the guidelines of academic writing and used correct grammar & spelling	4	8, 9, 10
15)	provided a good structure and layout of the assignment	1	1, 2, 3, 7
Total		20	

Back

Ability Distributions by Gender



Ability Distributions by Gender



Tutor gender analysis

	(1)	(2)	(3)
	Task Grades (Std)		
Mixed Team	0.204***	0.232***	0.199***
	(0.0614)	(0.0564)	(0.0625)
All Women	0.234**	0.274**	0.243*
	(0.109)	(0.109)	(0.123)
Female Tutor	0.111*	0.110*	0.0900
	(0.0647)	(0.0644)	(0.0690)
Mixed Team × Female Tutor	0.00716	-0.0145	-0.00724
	(0.0684)	(0.0634)	(0.0696)
All Women × Female Tutor	0.0778	0.0633	0.0447
	(0.112)	(0.120)	(0.116)
Best/Worst GPA Quint. (Uni)		✓	
Best/Worst GPA Quint. (HS)			✓
Observations	4,212	4,212	3,744

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Extended ability controls

$$Grade_{trg} = \beta_0 + \beta_1 Mixed_r + \beta_2 AllWomen_r + Task_t + Tut_g + \sum_{q=1}^{5}\sum_{p=1}^{q} \theta_{q,p} \mathbb{1}\left(AbilityQuintile_r^{Best} = q, AbilityQuintile_r^{Worst} = p\right) + \epsilon_{trg}$$

- $AbilityQuintile_r^{Best}$ - Ability quintiles for best in team
- $AbilityQuintile_r^{Worst}$ - Ability quintiles for worst in team
- Ability controls: High school GPA/University GPA

Extended ability controls

	(1)	(2)	(3)	(4)
	Task Grades (Std)			
Mixed Team	0.211***	0.227***	0.174***	0.200***
	(0.0452)	(0.0429)	(0.0518)	(0.0500)
All Women	0.280***	0.339***	0.284***	0.319***
	(0.0687)	(0.0699)	(0.0837)	(0.0863)
Uni. GPA Quint. Comb.		✓		✓
HS GPA Quint. Comb.			✓	✓
Mixed Team=All Women				
F Statistic	1.74	4.02	2.75	2.69
p-value	0.19	0.0472	0.101	0.105
Observations	4,212	4,212	3,744	3,744

1. Standard errors in parentheses, clustered on the small tutorial group level.
2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Back

Extended ability controls