UNIWERSYTET WARSZAWSKI
Wydział Nauk Ekonomicznych

Gender, beauty, and assessment of academic performance

Michał Krawczyk
University of Warsaw
August 2016

Implicit Associations Test

You will see a list of items. Your job is to categorize each item, one by one, as fast as you can. There are four categories. Here are the items that belong to each:

Category	Items
Male	Man, Son, Father, Boy, Uncle, Grandpa, Husband, Male
Female	Mother, Wife, Aunt, Woman, Girl, Female, Grandma, Daughter
Science	Astronomy, Math, Chemistry, Physics, Biology, Geology, Engineering
Arts	History, Arts, Humanities, English, Philosophy, Music, Literature

IAT-continued

- Place your hands on your knees.
- When the list shows up, start tapping:
  - Tap your left knee for each ``male'' word
  - Tap your right knee for each ``female'' word
- Go down the list as fast as you can, but avoiding mistakes
- When you're done, please look at the stopwatch on the screen to find out how much time it took you
- http://www.online-stopwatch.com/large-stopwatch/

Note the time when done

LEFT for MALE

RIGHT FOR FEMALE

Man, Uncle, Mother, Wife, Aunt, Woman, Girl, Female, Father, Grandma, Daughter, Son, Boy, Grandpa, Husband, Male

IAT-continued

- Prepare a piece of paper you will be able to hand in
- Those on my left-hand side
- Keep your hands on your knees.
- When the list shows up, start tapping:
  - Tap your left knee for each ``male'' or ``science'' word
  - Tap your right knee for each ``female'' or ``arts'' word
- Go down the list as fast as you can, but avoiding mistakes
- When done, write down the time on the piece of paper preceded by MS, e.g. MS: 18 or MS: 78
- Those on my right-hand side: please wait a moment

Only left-hand side!
Note the time when done

LEFT for MALE
or SCIENCE

RIGHT for FEMALE
or ARTS

Mother, Wife, Humanities, Grandpa, Aunt, Woman, Girl, Female, Father, Grandma, Astronomy, Literature, Chemistry, Engineering, Man, Uncle, Daughter, Son, Boy, Husband, History, Arts, Math, Male

IAT-continued

- Keep your hands on your knees.
- When the list shows up, starting tapping:
  - Tap your left knee for each ``male'' or ``arts" word
  - Tap your right knee for each ``female'' or ``science'' word
- Go down the list as fast as you can, but avoiding mistakes
- When done, write down the time it took you, e.g. MA: 47 or MA: 89

Note the time when done!

LEFT for MALE
or ARTS

RIGHT for FEMALE
or SCIENCE

Mother, Wife, Humanities, Grandpa, Aunt, Woman, Girl, Female, Father, Grandma, Astronomy, Literature, Chemistry, Engineering, Man, Uncle, Daughter, Son, Boy, Husband, History, Arts, Math, Male

IAT-continued

- Those on my right-hand side:
- Keep your hands on your knees.
- When the list shows up, starting tapping:
  - your left knee for each ``male'' or ``science'' word
  - your right knee for each ``female'' or ``arts'' word
- Go down the list as fast as you can, but avoiding mistakes
- When done, write down the time it took you
- Those on my left-hand side: pls hand in your slips

Those on the right-hand side! Note the time when done

LEFT for MALE
or SCIENCE

RIGHT for FEMALE
or ARTS

Mother, Wife, Humanities, Grandpa, Aunt, Woman, Girl, Female, Father, Grandma, Astronomy, Literature, Chemistry, Engineering, Man, Uncle, Daughter, Son, Boy, Husband, History, Arts, Math, Male

Discussion

- Gender-Science stereotype
- Categorization is typically easier (and hence faster) in the case of `` stereotype-congruent trials''—(female or arts) vs. (male or science)
- It is more difficult (more time needed, more mistakes) in the case of stereotype-incongruent trials—(female or science) vs. (male or arts)
- By comparing the speed of the two we can estimate the strength of individual bias

What is nice about IAT?

- It may reveal biases that subjects would not admit to
- Can be applied to various types of prejudice
- Seems to be internationally comparable
- Is somehow funny to take, so it is possible to reach high sample size

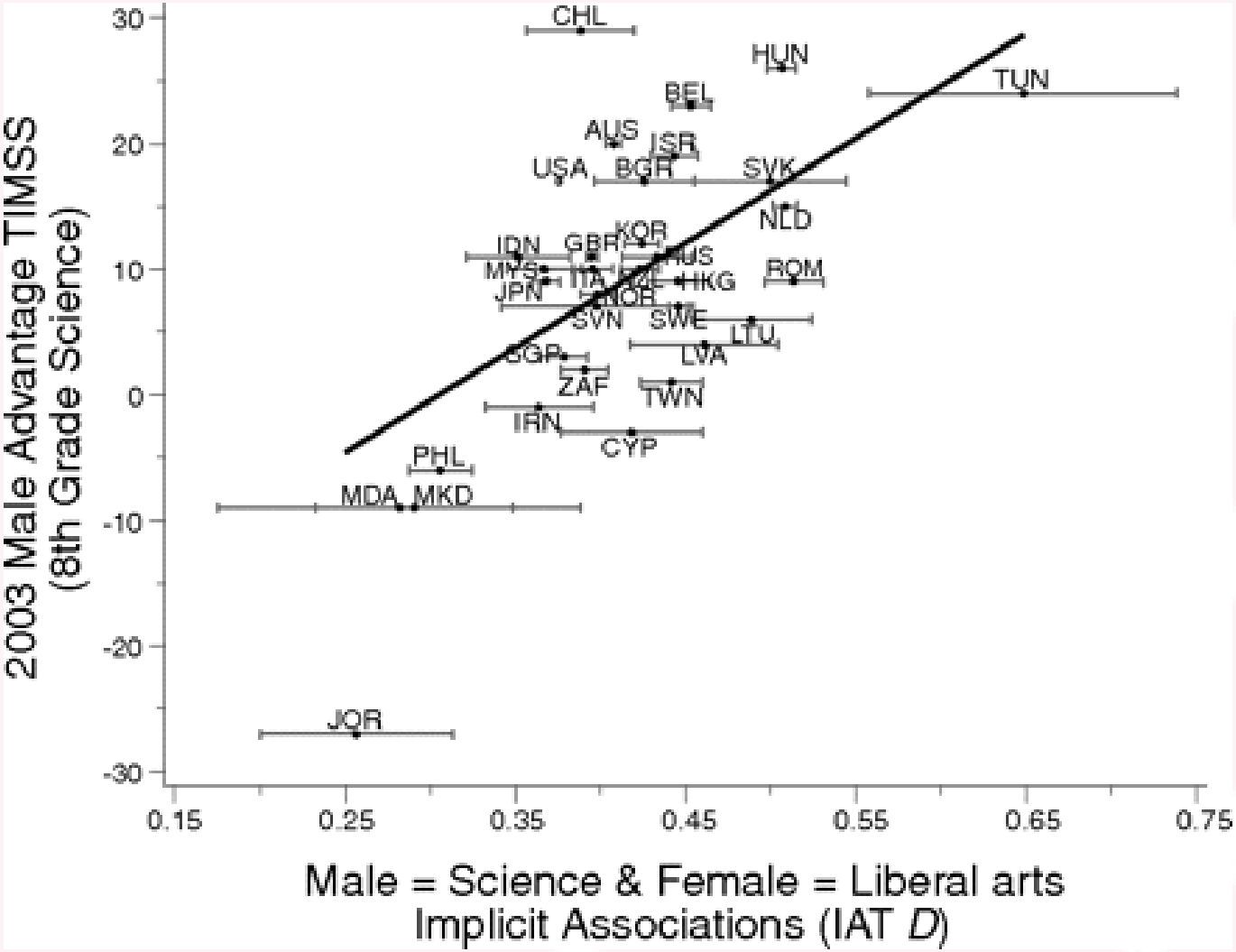
IAT on gender-science: findings

- Nosek et al. (PNAS, 2009) report data from more than 500k takers of the IAT
- Nearly 70% show the bias
- Perhaps even more interestingly, the authors link it to (8th-grade) gender gap in performance in science and math

Science and math gender performance gap

- Boys typically outperform girls in science (google TIMSS)
- This makes some believe there is a natural advantage
- For math, boys only do better in some nations
- Moreover, over the last decades, girls have improved a lot (Hyde et al., Science, 2008)
- This suggests socio-cultural factors play a role
- Could it be linked to stereotypes?

IAT correlates with science gender gap (r=.6)



Further observations

- the IAT only weakly related to self-reported stereotyping (r = 0.22).
- Self-reported stereotypes did not provide additional predictive validity
- Tentative conclusion: Implicit stereotypes and sex differences in science participation and performance are mutually reinforcing

Gender-science cliche can affect evaluation (Mechtenberg 2009)

- teachers get signals about the talent of their students,
- but these signals also depend on attitude
- they send messages (grades) to their students
- students choose efforts (reflected in PISA and TIMSS data)
- students choose college major and effort
- they enter labor market and are rewarded for their human capital

Gender-science cliche can affect evaluation (Mechtenberg 2009)

- teachers do not want to mislead students too much; if they think that the grade message will be taken very seriously, they want to send a true one
- a message that is likely to be incorrect will not be taken seriously (self-fulfilling prophecy)
- girls believe positive messages about talent in humanities, but do not trust bad messages
- as a result, girls tend to attend college and choose humanities
- boys who get a positive message about talent in math, choose math, others do not attend college

Predictios of the model by Mechtenberg (2009)

- boys outperforming girls in maths and sciences,
- boys having more top and more bottom achievers in maths and sciences than girls
- girls outperforming boys in reading literacy
- female graduates enrolling in university studies more often than male graduates
- the predominance of female students in arts and humanities at the university
- the predominance of male students in maths and sciences at the university
- the gender wage gap on the labour market for the highly educated

Gender and assessment: investigations using existing data

- Lavy (2008): boys do relatively well on “blind” tests as compared to non-blind tests in matriculation exams in Israel
- Similar results reported by Lindahl (2007), Angelo (2014), Falch and Naper (2013), Terrier (2014); Rauschenberg, (2014).
- Similar findings for tests vs. coursework (Kiss, 2013)
- Enzi (2015): among 5-6th graders, girls disadvantaged in math, but boys disadvantaged German
- Breda and Ly (2014): exactly opposite in French university entrance exams (consistent with Mechtenberg)

- The problem with most of comparisons using existing field data is that gender-format interactions may affect actual performance.
- Indeed, Cornwell et al. (2011): high grades of US female students were almost entirely explained when non-cognitive skills were controlled for.
- Girls may also underperform at exams due to greater test anxiety (Cassady and Johnson, 2008).

Gender and assessment: experiments

- Hinnerich et al. (2014) let randomly selected exams be anonymously re-graded. While overall scores were much worse, this did not interact significantly with gender.
- Van Ewijk (2010) and Sprietsma (2013) run field experiments letting teachers (in the Netherlands and Germany respectively) grade essays with randomly assigned fake names. This approach, aimed chiefly at finding out if minorities (e.g. Turks) are treated differently, typically led to no significant impact of gender being found.

Beauty premium at school?

- Loh (1993), Mocan and Tekin (2006), Umberson and Hughes (1987) speculated that pretty students get more attention from teachers
- Ritts et al. (1992): pretty students get higher grades and scores on standardized tests
- French et al. (2009): beauty correlates with GPA in high-school students

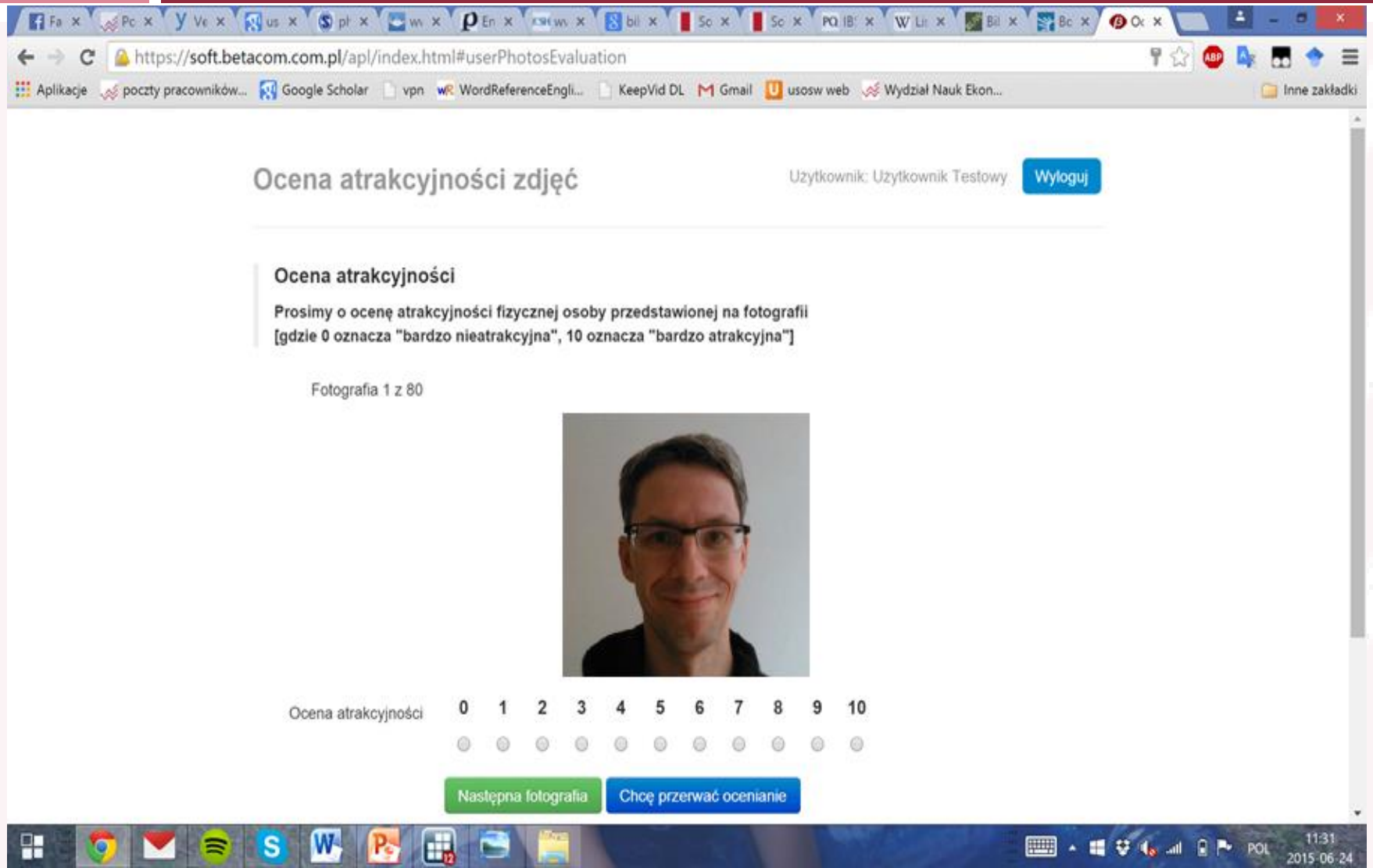
- Hernández-Julián and Peters (2015) collected data on grades, gender and rating of attractiveness of students at a US university
- They focused on how gender and beauty affected grades in online vs. traditional courses
- Finding: beauty premium much weaker for online courses
- (this supports the interpretation of unequal treatment rather than differences in performance merely correlated with beauty)

Hot or what? Do gender and beauty affect college grades? (Krawczyk, 2016)

- Each master and bachelor thesis in Poland is evaluated by advisor and referee
- The advisor, but typically not the referee, knows the student personally
- So gender more salient for the fomer
- She can also (perhaps inadvertently) take interpersonal attractiveness into account
- (Still, they evaluate exactly the same work)
- Will the advisor-referee gap depend on gender or beauty?

Data

- Over 15k graduates from several departments of the University of Warsaw
  - gender, advisor, referee, gpa, bachelor/master,
- Almost 3k of them also had photos rated on a 1-10 scale by one or more of our 50 raters
- 10 photos rated by all to check consistency. Cronbach alpha>.8


https://soft.betacom.com.pl/apl/index.html#userPhotosEvaluation
Aplikacje
poczty pracowników...
Google Scholar
vpn
WordReferenceEngli...
KeepVid DL
Gmail
ususw web
Wydział Nauk Ekon...
Inne zakładki
Ocena atrakcyjności zdjęć
Użytkownik: Użytkownik Testowy
Wyloguj
Ocena atrakcyjności
Prosimy o ocenę atrakcyjności fizycznej osoby przedstawionej na fotografii
[gdzie 0 oznacza "bardzo nieatrakcyjna", 10 oznacza "bardzo atrakcyjna"]
Fotografia 1 z 80
Ocena atrakcyjności
0 1 2 3 4 5 6 7 8 9 10
Następna fotografia
Chcę przerwać ocenianie
POL
11:31
2015-06-24

Gender does not matter

all advisors and referees			
	from advisor	from referee	AR GAP
females (*n*=11356)	4.709	4.619	.089 (s.d.=.426)
males (*n*=4303)	4.748	4.653	.093 (s.d.=.445)
male advisors and male referees only			
females (*n*=3343)	4.743	4.624	.118 (s.d.=.462)
males (*n*=2159)	4.784	4.674	.110 (s.d.=.451)

(2=failed 5=very good)

Gender still doesn't matter

	(1)	(2)	(3)
male student	.004	-.015	-.020
male advisor		.046***	.012
m. student* m. adv.		.020	.008
male referee		.005	-.013
m.student * m. ref		-.011	.003
master student		-.054***	.013
GPA		.030***	.033***
academic major dummies	NO	NO	YES

Beauty doesn’t matter either

	(4)	(5)
male student	.029	.020
standardized beauty	.003	-.002
male advisor	.050*	.043
m. student* m. adv.	-.025	-.03
male referee	.005	.001
m.student * m. ref	.005	-.050
master student	.021	.058*
GPA	.020	.038
academic major dummies	NO	YES
N	2607	2605

Discussion

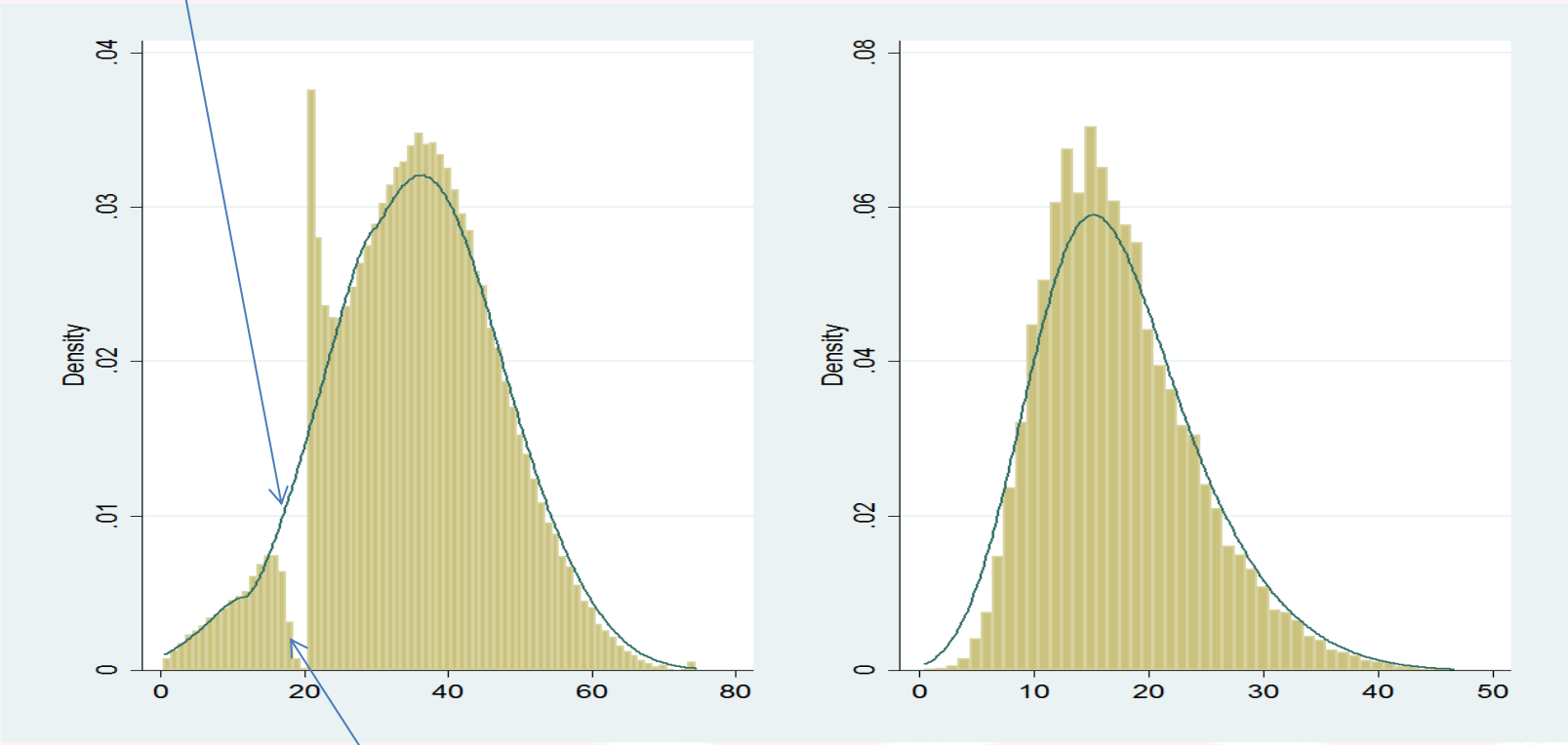
- Good news!
- Imperfect measures of attractiveness?
- Too little statistical power?
- Culture and institution-specific?
- Artifacts in previous studies?

Gender favoritism in high-school exams? (Krawczyk, 2016)

- Students take both written and oral exams
- Gender is less salient and beauty is not observable in written exams
- We may thus pursue previously discussed strategy of comparing gender gaps in these two cases

Additional idea: check who gets helped to pass?

How many ``should'' get 18 points



How many actually get it

Fractions of boys and girls who were *not* helped

Tests of equality of proportions in oral exams in Polish						
score	3,4, or 5		4 or 5		5	
	%	n	%	n	%	n
females	9.3%	45,198	9.4%	33,597	1.9%	18,663
males	11.0%	48,341	10.9%	35,473	2.3%	19,434
***p* value**	<.001		<.001		.003	

The same, this time for written exams

Tests of equality of proportions in written exams in Polish						
	18, 19, or 20		19 or 20		20	
	%	N	%	n	%	n
females	13.2%	23,738	4.8%	16,765	1.3%	8,875
males	13.3%	29,732	5.1%	20,879	1.3%	10,984
p value	.290		.222		.407	

Using the standard method

- To make the scores in oral and written exams comparable, they were both translated into ``normal scores''
- For instance, if given score in oral Polish exam was better than 90% of all oral Polish scores in this year (after random tie-breaking), it is assigned the value of 1.28, the ninth decile of the standard normal distribution
- We then run regressions explaining this score by gender and interaction gender*exam type
- We also run a logistic regression explaining pass/fail binary outcome

Results

Regression results, Polish language exams				
variable	ols1	ols2	logit1	logit2
oral exam	0.0838***	0.0569***	1.1143***	1.0074***
female	0.3752***	0.3764***	0.5425***	0.5835***
oral female	-0.1499***	-0.1516***	0.0324*	0.0449*
N	3 291 456	3 291 456	3 291 466	3 291 466
R^2	0.0591	0.0236		
legend: * p<0.05; ** p<0.01; *** p<0.001; OLS2 and logit2 include various demographic controls				

Conclusion

- Students' gender seems to have some (weak) effect on evaluation
- Bad female students are helped to pass more often than bad male students
- Mean female students may be treated a bit worse than mean male students
- Consistent with the stereotype of girls being more likable and boys being more competent